

Information Science and Technology Directorate
Defense Technical Information Center
Ft. Belvoir, VA 22060-6218

ADA478274

DTIC-E- TR-2008/05 March 2008
Information Science and Technology Directorate
Technical Report

CURRENT SEARCHING METHODOLOGY AND RETRIEVAL ISSUES: AN ASSESSMENT

by
Carl St. Clair Randall, Ph.D.



Approved for public release: distribution is unlimited

Information Science and Technology Directorate
Defense Technical Information Center
Ft. Belvoir, VA 22060-6218

DTIC-E- TR-2008/05 March 2008
Information Science and Technology Directorate
Technical Report

**CURRENT SEARCHING METHODOLOGY
AND RETRIEVAL
ISSUES: AN ASSESSMENT**
by
Carl St. Clair Randall, Ph.D.

Approved for public release: distribution is unlimited

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188								
<small>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</small> PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.												
1. REPORT DATE (DD-MM-YYYY) 17 March 2008		2. REPORT TYPE Final		3. DATES COVERED (From - To) March 2008								
4. TITLE AND SUBTITLE CURRENT SEARCHING METHODOLOGY AND RETRIEVAL ISSUES: AN ASSESSMENT				5a. CONTRACT NUMBER								
				5b. GRANT NUMBER								
				5c. PROGRAM ELEMENT NUMBER								
6. AUTHOR(S) Carl St. Clair Randall, Ph.D.				5d. PROJECT NUMBER								
				5e. TASK NUMBER								
				5f. WORK UNIT NUMBER								
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Information Science and Technology Directorate Defense Technical Information Center Ft. Belvoir, VA 22060-6218				8. PERFORMING ORGANIZATION REPORT NUMBER DTIC-E- TR-2008/05								
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Information Science and Technology Directorate Defense Technical Information Center Ft. Belvoir, VA 22060-6218				10. SPONSOR/MONITOR'S ACRONYM(S)								
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)								
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release: distribution is unlimited												
13. SUPPLEMENTARY NOTES												
14. ABSTRACT The study examined the current state of searching and the preferred method as indicated by the participants. Their views were analyzed and presented in several categories, they are: preferred method of searching; status of searching methodology and its future; search systems performances and the ability to effectively measure these systems; improvements needed in search systems; future role of catalogers and indexers; and searching in the future.												
15. SUBJECT TERMS Search Methodology, Full-Text Searching, Metadata Searching, Search Engines, Search Systems, Search Systems Performances, Retrieval Issues, Barriers to Searching, Unstructured Data, Structured Data, Future Searching, Catalogers and Indexers												
16. SECURITY CLASSIFICATION OF: <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%; padding: 2px;">a. REPORT</td> <td style="width: 33%; padding: 2px;">b. ABSTRACT</td> <td style="width: 33%; padding: 2px;">c. THIS PAGE</td> </tr> <tr> <td style="height: 30px;"></td> <td></td> <td></td> </tr> </table>			a. REPORT	b. ABSTRACT	c. THIS PAGE				17. LIMITATION OF ABSTRACT		18. NUMBER OF PAGES	
a. REPORT	b. ABSTRACT	c. THIS PAGE										
					19a. NAME OF RESPONSIBLE PERSON Carl St. Clair Randall, Ph.D.							
					19b. TELEPHONE NUMBER (Include area code) 703 767-9105							

ACKNOWLEDGMENTS

I am grateful to all participants who volunteered to be a part of this research study report. Each participant was gracious enough to acknowledge the importance of the research by the interest that was demonstrated in the interviews and follow up conversations.

I am personally indebted to my colleagues who provided insight and comments throughout the research experience. I would like to thank Carol Jacobsen, Gopi Nair, Roberta Schoen, Grant Clark, Steve Markheim, Marjorie Powell, LaDona Kirkland, Nancy Wright, Christina Pikas and Marcia Kerchner for their generosity in draft reviews, comments and suggestions to improve the final paper. Also, I would like to thank Cherie Lee, Warren Carman, and the Micrographics staff who played a significant role in producing the final product.

TABLE OF CONTENTS

Acknowledgments.....	i
Executive Summary.....	6
Introduction.....	8
Background.....	8
Purpose of Study.....	9
Report on current views on searching	
Assess the future of searching	
Methodology.....	9
Interview Questionnaire	
Sampling Population	
Organizational Grouping	
Content Analysis	
Review of Literature.....	12
Improving Searching Methodology	
Retrieval Issues and Barriers to Searching	
Information Retrieval	
Measurement and Performance Evaluation	
Relevance Ranking	
Search Engine Capabilities	
Desired Improvements to Searching	
Future in Searching	
Role of Catalogers and Indexers	
Data Analysis.....	31
Summarized Responses by Group	
Conclusion.....	72

TABLE OF CONTENTS

(Continued)

Appendix A: Table Summary.....	84
Appendix B: Interview Questionnaire.....	226
References.....	231

EXECUTIVE SUMMARY

HIGHLIGHTS

Current search methodologies have a direct impact on the fundamental retrieval issues that information seekers encounter in their use of the vast number of search systems on the web today. Both novice and expert searches face a number of challenges in their web searches, such as relevant search results, quantity and quality of hits, barriers to effective searching, and the ever changing volume of data that is available. These challenges can be intimidating and discouraging to the occasional information seeker who may be looking for an answer to a question but may not know where to begin. The more experienced seekers of information also face challenges in obtaining the answers to questions or finding the available data about their subject.

This study examined some of these issues by reporting on the literature reviewed about the subject. Interviews were also conducted with a cross section of information professionals. Their responses were analyzed and presented in the report.

The two primary methods of searching that are used by search engines are discussed. They are: full text searching, i.e., the searching of unstructured data, and metadata searching, i.e., the searching of structured data. In the latter case, there is a controlled vocabulary or thesaurus provided. Hybrid search systems are also found among search engines; however, it is the popularity of full text searching that has changed the road map to information access.

The methodology used for this study was to conduct an extensive review of the current literature on the subject to access the state-of-the-art. Secondly, a selected group of information science professionals were interviewed from a cross section of government agencies, educational institutions, and private sector organizations. An interview questionnaire was developed that comprised 26 questions and statements to solicit the personal views of the participants. The views expressed are those of the participants and are not the positions held by their respective organizations or institutions.

Twenty nine organizations and institutions were selected for inclusion in the study. There were 48 participants grouped into five sub groups that best describe their organizational and institutional missions and goals. The 26 questions and statements were grouped into seven categories. Each subgroup was evaluated against the categories to form 35 tables that illustrate the participants' responses. The tables are provided in appendix A of the study report.

The participants provided various reasons for their preferred method of searching. A few stated that full-text searching was their preferred method. The primary reason was the belief that it is easier and faster to conduct a full-text search. There were a few participants whose search preference was metadata searching; however, the majority of participants used both methods.

Participants were asked to express their views on the status of searching methodology, and its future. Flexibility in conducting searches was emphasized. The view is held by some that having an access system to accommodate both full-text and metadata searching would be ideal. Participants also believed that there is an ongoing challenge for content providers to develop search systems that meet the needs of specific communities of practice.

The study also examined search systems performances and the ability to effectively measure these systems. The overall responses support the need for improvement in their ease of use. There was support for improvement in search tips and help guides. Improvements in interface design and usability to promote more seamless search systems was strongly recommended.

Several fundamental flaws were identified by participants and were also supported in the reviewed literature on the way search system performances are measured. The current literature identifies shortcomings with the vast majority of broad base search systems such as Google, Yahoo, and MSN. Their ease of use comes with a price that information seekers find unacceptable.

The study also addressed what participants viewed as improvements needed for search and retrieval effectiveness, and some of the barriers to overcome in order to improve the information seekers' experience.

The next area that was addressed in the study focused on the future role of catalogers and indexers and the overall role of online catalogs. The study then examined the future role of online catalogs in light of other discovery tools.

The study also examined searching in the future. What will web searching in the future provide that is currently lacking from one's search experience? While there are differing views on the future of searching, the consensus is that technological advancements in search systems and improvement in information harvesting across multiple databases on a global platform will play an integral part in search and discovery.

INTRODUCTION

Over the past 40 years there have been substantial changes in searching capabilities and retrieval effectiveness. Online searching has increased information seekers' access to information, leading to a shift in the role of traditional library and information centers. Seekers of information are now equipped with the tools to conduct independent searching for information. A decade ago the information was only available in a static state, i.e., on the selves of libraries, storage centers of organizations and institutions, or perhaps in archival storage.

The power of the Internet has served as a gateway to information access across geographic boundaries and institutional fiefdoms. While barriers do exist to information access, the availability of information and the speed of access today are quantum leaps ahead of search and retrieval capabilities prior to the Internet. Minor inconveniences in information sharing whether regional, national, or global will be improved as more information is made available, information seekers demand more access to information, and there is the inevitable improved openness of information.

BACKGROUND

Searching methodologies and retrieval effectiveness have changed the scope of access to information. Early online access to database such as ERIC marked a shift from the card catalog as the gateway to information to a more robust way of determining available information on a specific subject. The card catalog became an online one. More seekers of information now have access to the tools to conduct inquiries. As more powerful searching capabilities evolved, the methods of searching grew. This has led to greater access to documents in their entirety. The two fundamental searching methodologies applied today, full-text searching (searching of unstructured data) and metadata searching (searching of structured data), provide information seekers with more flexibility in searching. There are systems available that adopt elements of both methodologies. The popularity of full-text searching has changed the roadmap to information access. This is clearly obvious with the advent of Google and Yahoo as two of the dominant providers of information access. Information seekers' demands for quick and easy access to information often lead to vast amounts of unrelated or irrelevant information on a particular subject search; however, the recipient may not be concerned with the vast number of hits if the answer to his/her question or need is met. On the other hand, information seekers' willingness, or lack of, to learn the multiple search engines' capabilities may diminish their search results.

The issues surrounding metadata and full text searching are addressed in this study. A review of the literature and interviews conducted with professionals from the information science discipline provides insight into the status of searching and retrieval.

PURPOSE OF STUDY

The purpose of the study was to assess the status of searching methodologies. Answers were ascertained for the following questions: What are some of the current and desired searching capabilities? What are some of the limitations that need to be addressed in order for information seekers to obtain what they need? What are the preferred methods of searching and the rationale for these decisions? What will search systems in the future provide that is currently not available?

METHODOLOGY

In an attempt to address questions pertaining to searching methodologies, a review of the literature on the subject was conducted. Also, Information Science professionals were identified from a wide variety of organizations for inclusion in the study. An interview questionnaire was developed as the tool for gathering individuals' thoughts and views. The questionnaire was comprised of 26 questions and statements from which participants' responses were sought. The questionnaire was administered in three forms: in-person interview; e-mail, and telephone contact.

Twenty nine organizations and institutions were selected for inclusion in the study. Organizations were grouped into five subgroups that best describe their mission and goals. They were:

- CENDI member agencies (an interagency working group of senior scientific and technical information (STI) managers from federal agencies)
- Department of Defense (DOD) Organizations and DOD Contractors (library professionals)
- University Information Science and Computer Science Department Professors
- Information Science Organizations
- Other Libraries

Participating organizations within sub-groupings:

CENDI:

- Defense Technical Information Center (DOD)
- Government Printing Office
- Library of Congress
- NASA Scientific and Technical Information Program
- National Agricultural Library (Department of Agriculture)
- National Archives and Records Administration
- National Library of Medicine (Department of Health and Human Services)
- Office of Scientific and Technical Information (Department of Energy)
- USGS/Biological Resources Discipline (Department of Interior)

DOD Organizations and Contractors

- Air Force Research Laboratory
- Chemical and Biological Information Analysis Center (CBIAC). The name has been changed since the interview was conducted. The new name is: Chemical, Biological, Radiological and Nuclear Defense Information Analysis Center (CBRNIAC)
- Johns Hopkins University, Applied Physics Laboratory
- Lackland Air Force Base
- MITRE Corporation
- Naval Research Laboratory
- Pentagon Library
- Picatinny Arsenal
- Redstone Scientific Information Center (RSIC)

University Information Science and Computer Science Departments

- Old Dominion University
- San Jose State University
- Syracuse University
- University of North Carolina, Chapel Hill

Information Science Organizations

- Access Innovation Incorporated
- Information International Associates Incorporated
- National Commission on Libraries and Information Science (NCLIS)
- National Federation of Advanced Information Services (NFAIS)
- Southeastern Library Network

Other Libraries

- Catholic University
- US Senate Library

The 48 participants represented 29 organizations and agencies. Participants included: Information Science Professionals (senior managers, technical information specialists, and librarians) from the Scientific and Technical Information (STI) community within the federal government, Reference Librarians and other information providers from the university community, University Professors from information science and computer science departments from several universities, Professionals from various information science organizations and

companies, and Information Professionals from non-CENDI Federal agencies and Government-supported organizations.

Questionnaire responses from the 48 participants were grouped into seven broad categories (see Appendix A).

The categories were:

- Preferred Method of Searching
- Searching Methodology...Full Text, Metadata, Other
- Limitations in Full Text and Metadata Searching
- Search Systems Performance...Measures
- Improvements in Retrieval Effectiveness
- Future Role of Catalogers and Indexers
- Improving Search Results...Role of Metadata and Full Text.

The data was analyzed using content analysis. The seven broad categories served as a way to group similar and related questions and statements.

REVIEW OF THE LITERATURE

Improving Searching Methodology

Changes in searching methodology can be viewed as a systematic approach to an iterative process by improving search and discovery. Search improvements necessitate an assessment of the search system and an understanding of the information seekers' behavior. Maybury's (2005) presentation on "Making Search and Discovery Work" addresses both the barriers and a range of potential solutions to search improvement methodology. He suggested that a technology assessment be conducted in which the search system capabilities and activities be analyzed. In addition, the barriers to retrieval must be understood in order to improve searching methodology. An analysis of the tasks, corpus, user/usage/usability with an understanding of the information seeking behavior of users, their query intent versus query results, the adequacy of the search, and their navigational capability goes a long way in realizing improvements. The author's roadmap for search improvement methodology includes: *a shift in focus from defining metadata to analyzing usage, engaging vendors, infusing practice with systems engineering rigor, and optimizing search locally.*

Government websites (estimated to be in excess of 17,000) include a large portion that lack search interfaces, making searching a challenge. Hawking and Thomas (2005) proposed a hybrid approach to access, whereby a combination of distributed and centralized techniques is applied. The authors advocate distributed methods where network bandwidth is limited or expensive. Servers with search interfaces would be candidates for metasearch, and the others would be crawled. Hawking and Thomas acknowledged, however, that a hybrid centralized/distributed replacement for FirstGov would be highly unlikely, due to the low cost and the wide availability of bandwidth.

Retrieval Issues and Barriers to Searching

We are near the end of the second decade since the first Internet search engines were developed. A fundamental problem that information seekers still face is how to retrieve the information that is sought. Search engines are still trying to figure out how to improve the accuracy of responses to questions by information seekers. One approach is to combine searching with new technology. The fundamental issue of improving searching capabilities by removing barriers to retrieval effectiveness remains. The high speed Internet has helped to lower the barrier as more information seekers gain access.

In 1986, Borgman reported the difficulties in the use of online catalogs. The primary reason noted was designer's lack of understanding of information seekers' behavior. Search systems were designed to accommodate the skilled intermediaries and not the end-users. Some 10 years later, Borgman revisited the issue of online searching and found that little had changed. The author points to the fact that studies on "information seeking" have shown that information seekers formulate their questions in stages, and eventually articulate a query. "A search may be conducted over a number of sessions with different information technologies and sources, both online and offline, picking and choosing from multiple options to answer a question or explore

an issue. The design, however, of most operational online catalogs assumes that information seekers formulate a query that represents a fixed goal for the search and that each search is independent,” Borgman (1986). The author concludes that in spite of improvements in user interface designs of online catalogs, information seekers still find them difficult to use. Improvements have more surface features rather than core functionality.

Information seekers’ experiences in searching can be viewed from both a technological and human perspective. While the ability to retrieve relevant or accurate information may be attributed to search engine capabilities and the human inputting of data, there are fundamental issues with regard to the searcher’s or information seeker’s behavior that should be understood in order to facilitate successful outcomes. Carol Tenopir noted in her presentation at the 2005 Search Engine Meeting in Boston that there are probably 200 good studies over the last decade that have addressed user searching behavior. These studies analyzed usage logs, interviews, surveys, critical incident, and users in controlled settings. The information gathered from these studies may be useful; however, there remain fundamental flaws in these data gathering techniques. Tenopir identified clear distinctions between student and expert search experiences. Students select Internet search engines versus formal electronic sources, (such as online catalogs) as their first choice in searching for information. Their focus is on simplicity and speed. They value multitasking. On the other hand, expert searchers do both browsing and searching. Their usage pattern varies by subject. Collectively, information seekers use print and electronic sources. They tend to print those resources in which more time will be spent in reviewing.

Scholars have consistently emphasized in their research studies the importance of “best practices” in designing user interfaces. It is perceived that designers of these interfaces are faced with the challenge of appeasing the expert searcher while accommodating the novice users who may demonstrate little or no desire to learn the rules (understanding each search engine’s architecture and algorithms). The lack of understanding leads to frustration and poor search results (Resnick and Vaughan 2006).

The early user interfaces were primarily designed to facilitate the needs of expert searchers in accessing large corporate databases, library, and government information (Rappaport 2002). Intermediaries could input Boolean queries to obtain relevant information for the information seeker (user) who may or may not have been the searcher. Information Retrieval (IR) eventually became more accessible to the larger segment of the population; however, the complexity of these IR systems proved too difficult for the novice user. Novice information seekers (users) accessing public libraries had the added benefit of obtaining assistance from expert librarians. In contrast, the information seekers had to rely on their own capabilities when conducting searches from a home computer. *Search Engine Watch* in 2000 estimated that 18% of users surveyed had difficulty finding what they were looking for on the Web, while 67% stated that they were frustrated while searching. Sullivan (2000) expressed similar results.

Resnick and Vaughan (2006) noted that “best practice” suggests design superiority over other ideas. The designers should first identify the audience and then determine their needs for the system. The authors further noted that system design should be treated as an ongoing and iterative process by consistently looking for improvements and fine tuning as information

seekers' needs and demands change. Resnick and Vaughan presented a summary and analysis from several researchers and user interface designers' views on "best practices," at the Conference on Human Factors in Computing Systems (2003). The search design best practices were divided into **five categories**:

1. Structure of the database

The authors believe that an understanding of the nature of the database structure is essential prior to designing an effective search system. The parameters of the search system must be clearly understood. Will the search engine access the entire Internet, or is it limited to a domain or cluster of domains, such as, all medical sites or all Department of Defense Laboratories? The authors further note that such differentiation is important since search systems that are all inclusive are limited in the assumptions that they can make about their content. In contrast, those search systems that have limited or defined domains have more control over the content.

The design of the user interface should also be influenced by the diversity of the content within the database. Resnick and Vaughan referenced the Davis' (2006) study on "improving internet interaction" where the researcher illustrates that in cases where search systems parse a single site, the diversity of information may vary widely. The Digital Library for Earth Science Education (DLESE) is an example of a consistently structured database that has a controlled vocabulary. In contrast, the AOL e-commerce database is more diverse (Gremett 2006), increasing the chances for more false hits and limiting the use of a controlled vocabulary and metadata.

2. Matching Algorithms

Algorithms are used to parse the database and match queries to content. Resnick and Vaughan noted that even when a database is comprehensive and organized, the absence of an effective algorithm to match queries to specific content leads to unsuccessful search results. Query expansion by adding synonyms and other related words is advocated as a means of minimizing that concern. The net result is increased hits. Also, query contraction is a way to remove terms with multiple meanings to improve the number of relevant matches in the search result. The application of natural language processing to queries is viewed as another way to improve matching (Zhou and Zhang, 2003).

In summary, Resnick and Vaughan advocate the use of domain-specific dictionaries and thesauri, spell checking of terms for queries, and document level expansion for algorithms matching.

3. User Content and Task Requirement

The search methodology applied by an information seeker will vary with the search task, the searcher's knowledge of the domain being searched, accessibility to the knowledge base, and perhaps, the available time to conduct an inquiry. Hearst et. al. (2002) identified four search types that an information seeker may apply when conducting a search. They are:

- direct search (a search for a specific item or fact, e.g., the year the United States became an independent country)
- comparison search (a search for information about multiple items in order to compare, e.g., cordless phone brands)
- informal browsing (a search for general information on a topic, e.g., starting a flower garden)
- text mining and analysis (a comprehensive search for information on a specific topic, e.g., non-smoking women with lung cancer).

Search strategies may also be viewed as top-down searching versus bottom-up searching. In Thatcher's 2000 presentation to the Human Factors and Ergonomics Society, he noted that information seekers may conduct an inquiry using general search terms and may subsequently introduce more specific words or terms from the initial result to further explore their findings. The opposite is also true, where a bottom-up approach may be undertaken. The inquirer may begin a search by using specific keywords and expand the search to retrieve the appropriate number of "hits" desired. An information seeker may move from a searching to a browsing mode and vice versa, or may use any, or all, of the search methods mentioned above. The information seekers' domain knowledge will ultimately dictate the search strategy used. Jefferson and Nagy (2002) reported that an information seeker and the search system apply the same term in only 10-20% of the time.

4. Interface between the information seeker and search system

Resnick and Vaughan identify the fourth search design best practice as the interface between the information seeker and the search system. The authors divided the interface into three groups. There is an input interface. The size of the input or search box will dictate the amount of data an information seeker will provide in conducting a search. Bandos and Resnick (2004) found that more effective queries were realized when interfaces provided brief guidance on search syntax and semantics. Search hints located near the search query box also proved to be beneficial in conducting a search. The second group, the output interface, contains the fields that search designers perceive as being most important. A fundamental issue is deciding how many results to include from a search. Results divided into categories or folders are useful tools for the information seeker. It makes the task more manageable when analyzing search results. Finally, there is iterative searching, where an information seeker gathers content from previous queries, modifies the queries, and seeks more information on the subject matter through further searching.

5. Emergence of hardware and bandwidth challenges with mobile devices

A different approach to search system design is required when access to data is obtained through mobile devices with limited bandwidth and small screens. In Jones, Buchanan, and Thimbleby's 2003 study on improving Web search access on small screens, the authors advocated versions of content specifically for viewing on small screens. Resnick and Vaughan summarized "best

practices for searching on mobile devices” to include: the design of alternate versions of content, scrolling versus switching between pages, and vertical rather than horizontal scrolling.

At the 10th Search Engine Meeting in Boston, Massachusetts, (2005), Hans Henseler stated that high precision and recall are necessary in the Law Enforcement/Intelligence discipline. In this field, the information seeker cannot afford to miss any relevant documents, so a 100% recall is necessary. However, technology alone cannot adequately increase precision when 100% recall is needed. When the information seeker is allowed to determine what is relevant in the search experience, precision will improve.

At the Search Engine Meeting in 2006, Tony Gentle’s presentation “A Healthy Perspective on Search Behavior” emphasized that there is a difference between searching for something and researching something. He noted that this is especially true in the health professional field. The key is to connect consumer with medical vocabularies to provide a medically-guided search.

Information Retrieval

What is information retrieval? A commonly used definition is the searching for information that resides in a document or documents, the searching for metadata, or searching within databases. These databases may be relational, “stand-alone,” or hyper textually-networked as the Web.

Belkin and Croft (1992) conducted a study to examine the relationship between information filtering and information retrieval. The authors concluded that they are “two sides of the same coin” with the ultimate goal of helping information seekers’ find the answers to their questions or needs.

For the purpose of this study, information retrieval is defined as the ability to access data in multiple formats (documents and multimedia) from search systems to satisfy an information seeker’s needs. Search results may be favorable or unfavorable.

Belkin and Croft (1992) identified the three early primary information retrieval models as Boolean, vector space and probabilistic. Boolean model is based on “the exact match” principle while vector and probabilistic are based on “the best match” principle. A fundamental shortcoming with Boolean is its inability to factor in relevance ranking of the retrieved documents set (Belkin 1987).

The vector space model treats texts and queries as vectors in a multidimensional space. The more similar a vector representing a text is to a query, the more likely the text is relevant to that query. Terms can be weighted to account for levels of importance. They are computed based on statistical distributions of terms in the database and text (Salton, 1983). In Salton’s 1975 research on vector space model for automatic indexing, the author noted that in document retrieval it appears that “the best indexing (property) space is one where each entity lies as far away from the other as possible. The value of the indexing system is a function of the density of the object space. Retrieval performance may correlate inversely with space density.”

The third retrieval model discussed by Belkin and Croft is probabilistic. The authors also view this model as based on “best match” principle. It assumes that there are several sources of evidence that could be used to estimate the probability of relevance of a text to a query, such as the statistical distribution of terms in a database.

In the 1980s, information retrieval systems were based on a “best-match” principle. The basis of this premise was that an information seeker’s request for information through a query or set of index terms would derive the text that most closely matches those search terms. Davies (1978) explained, “best-match principle depends upon the assumption of equivalence between expression of need and document text in that it treats the representation of need as a representation of the document ideal for resolving that need.” Belkin, Oddy, and Brooks (1982) also supported the best-match principle theory.

Have we improved retrieval systems effectiveness? The popularity of full-text searching (Google, Yahoo, etc.) has increased information seekers’ (with various abilities) access to a wide array of information that a decade or so ago would have been accessible only to a limited number of searchers. This increase in popularity has also brought a false sense of hope to the millions who believe that all information is free and can be accessed on the Internet. To most users, the ease of access perhaps outweighs the vast number of “hits” with low precision. In contrast, metadata searching minimizes this problem through the use of a controlled vocabulary. The ambiguities may be less, but the search may produce low recall by failing to identify and retrieve documents relevant to the query. The quality of the indexing goes a long way in determining the effectiveness of one’s search results.

How can the performance of full-text-searching be improved? Improved query tools are one way to achieve success. They include: Boolean queries, phrase searches, proximity searches, and quality keywords assigned to the document.

In Blair and Maron’s (1985) classic study of a full-text document retrieval system containing some 350,000 pages of text, the authors noted that the search system retrieved only 20 percent relevant or useful documents. An evaluation was conducted on IBM’s full-text retrieval system, Storage and Information Retrieval System (STARS). Blair and Maron concluded that full-text retrieval systems applied to large databases are not likely to perform well. Improvements in retrieval effectiveness may be realized if the information seeker rather than an intermediary does the search. The information seeker would do both the query formulation and modifications. Another reason for low recall is due to the difficulty in retrieving documents by subject. The authors concluded that early studies that demonstrated higher relevancy were based on small databases. These studies were also designed to show that full-text searching was competitive with searching based on manually assigned index terms.

Measurement and Performance Evaluation

The traditional way of managing complex systems is to divide them into subsets or subgroups, manage them as separate entities (evaluating through performance measures), and assuming that if each subset or subgroup works well, the whole system will also work well. Ackoff (1993)

views this approach with skepticism. He argues that while the performance of each subset or subgroup may improve in its performance, the system as a whole may not necessarily respond in a positive manner.

Nicholson (2004) advocated the application of a holistic evaluation, whereby, “the individual subset or subgroup can be combined to produce something beyond the sum of the individual subset or subgroup” (Wilbur, 2003). Nicholson noted that for measurement and evaluation of a system, “a more thorough knowledge and understanding can be attained by combining different measures, than if one were to conduct those measures separately.” Ackoff (1993) reported that the entire system must be evaluated to fully understand the effects of changes to any portion or subset of the system.

The wide variety of electronic information resources available to information seekers presents a challenge in measuring performance or success of search systems (Ma 2002). Information seekers now conduct their own inquiries, a role that traditionally was performed by information professionals. In-person consultation with librarians has given way to individuals independently accessing resources through the vast discovery tools that are now available through the Internet. Scholars and students can readily access remotely the resources that libraries have made available electronically. How does one evaluate performance of these resource providers (search systems) when the access and retrieval of data is dictated by the information seeker not the information provider?

The traditional way of measuring search systems’ performance was by determining precision and recall ratio for a specific search system. Earlier studies such as the MEDLARS search system for medical literature at the National Library of Medicine used such an approach. Lancaster (1969) noted that MEDLARS was not an end user searching system since the information seeker had to submit search requests to the library where they were administered. The results were sent to the information seeker for an analysis of the precision and recall. Current thinking supports a holistic approach to performance evaluation. The information seeker plays a relevant role in the development of the interface with the search system. A system evaluation should include: usability testing and assessments; user satisfaction surveys; search logs; reports of system response time and downtime; success of information seeker queries; and the frequently used search terms that are excluded from the search systems controlled vocabulary.

Kerchner (2006) supported methodologies for improving search experience that include both the information seeking task, the quality of outcome, and information value to the customer. This view is also supported by Nicholson (2004) who emphasized the importance of a holistic view to performance measurement. Nicholson and Kerchner noted that any effort to improve the information seeker search experience goes far beyond high precision and high recall search results. Performance measure must also include the total search experience. An evaluation must be conducted to determine if the search system helped the information seeker solve his search task. When does the information seeker search experience begin? Kerchner believes that the experience starts upon entering key words in a search box and goes through the search system feedback with search results. The information seeker’s assessment of the usefulness of the information retrieved must be included in the evaluation of a search success.

Kerchner noted that the traditional way of addressing low precision or low recall was through a **metadata solution approach**, involving adding of topical tags to content objects based on controlled vocabularies or perhaps through the replacement of the search engine. The author pointed to the high cost of maintaining taxonomies, their inconsistencies, and the fact that the taxonomy is the view of an individual or a group of individuals. Topical metadata is often implemented without much knowledge or understanding of the types of queries or the information seeker's search behavior. Another approach to improving search results is to replace a search engine. Kerchner warns that there is no guarantee for success without a true understanding of the barriers to an information seeker's search experience. The author recommends fine tuning the search process for improved results.

Kerchner identified six approaches for improving information retrieval.

- **Document engineering** which involves adding terms that are good discriminators and also reflect commonly entered search terms to content to improve retrieval effectiveness and the establishment of content quality standards.
- **Query enhancement** in which results from the information seeker's queries are reviewed and new terms are added to resubmitted queries to enhance search results.
- **Search improvement** can be achieved by intercepting popular queries and returning preconfigured results. Also, adjusting the search engine's parameters, such as placing more weight on specific metadata tags, can improve relevancy.
- **Results ranking improvement** takes place when search results are programmatically re-ranked prior to the information seeker viewing the results...this may include the re-ranking of multiple search results.
- **Categorization** in which large sets of results are grouped into subsets can enhance findings.
- **Summarization** in which passage-based summaries and highlighted search terms appear in the summary and content of the retrieved results.

While researchers have advocated a wide array of methods to improve information retrieval, the fundamental question remains about the cost to organizations when the information sought is not found or is missing from the sources that are searched. Information seekers are already faced with the challenge of filtering too much information that is located in multiple sources (databases and repositories) both within their organizations and as open access data. The lack of single access points or unified ones increases the risk of decision making based on incomplete information. Such decisions could lead to manufacturing failures, waste, slow response time, and poor standards or work output.

Susan Feldman (2004) identified a high cost associated with not finding the information sought. The author noted that information disasters are a growing threat, with missing or incomplete information plaguing project outcomes. The International Data Corporation (IDC), in 2001, looked at the cost to organizations when information critical to decision making is not found. They concluded that approximately 50% of web searches are abandoned by searchers. Feldman noted that in studies conducted by IDC, Association for Information and Image Management (AIIM), the Ford Motor Company, and the Working Council of CIO's, the following conclusions were made: knowledge workers spend 15-35% of their time searching for information; the success rate in finding what was sought was only 50%; and only 40% of corporate users found the information that was sought on their respective intranets. The author further noted that in an IDC 2001 study in which an attempt was made to quantify enterprise search, only 21% of the respondents found the information that was needed 85-100% of the time. There is an economic cost when knowledge workers are required to recreate or rewrite information that cannot be located within their organizations' databases. Feldman noted that '**information disaster**' occurs when there is an inability to connect the right information to the right people at the right time. The author further noted that since information is used in the context of what the decision maker is doing, it is critical that access to the right information is available when it is needed. "There must be assurance that access is guaranteed, easy, fast and reliable," (Feldman, 2001).

Relevance Ranking

The evaluation of information retrieval systems has been based on the relevance of the documents found in a particular search. Traditionally, the effectiveness of a search experience was measured by calculating the recall and precision values. Jacso (2006) reported that in "sample test using the same databases but on different host, there were significant differences in the relevant ranked result list for functionally identical queries." The author concluded that there was "a lack of consensus among search systems when determining the topical relevance of the same documents or document surrogates within the same database context." An explanation provided for these differences suggested that new records added to a database may change the ranking test results since perfect synchronization is difficult to achieve. Also, adjustments to search systems' algorithms could lead to differences in ranking positions for documents retrieved.

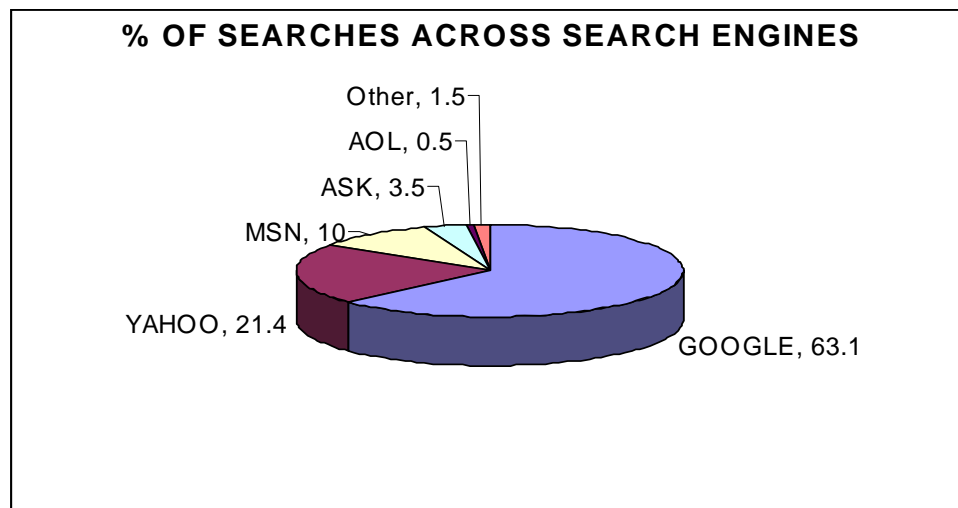
The characteristic patterns of information seekers' behavior have been addressed in large scale studies where Web logs were analyzed. Studies by Silverstein et al (1999) and Spink (2001) provided insights regarding behavioral patterns. These behavioral patterns influence relevance ranking. Results from these studies showed that information seekers seldom looked beyond the first screen; few used Boolean, proximity, positional operators, or even attempted to reformulate their queries. Also, the majority of information seekers did not use quotation marks for phrase searching.

Search Engine Capabilities

A frequent complaint among information seekers searching the web for answers to their questions is how to manage the vast number of hits received. This is more frequently

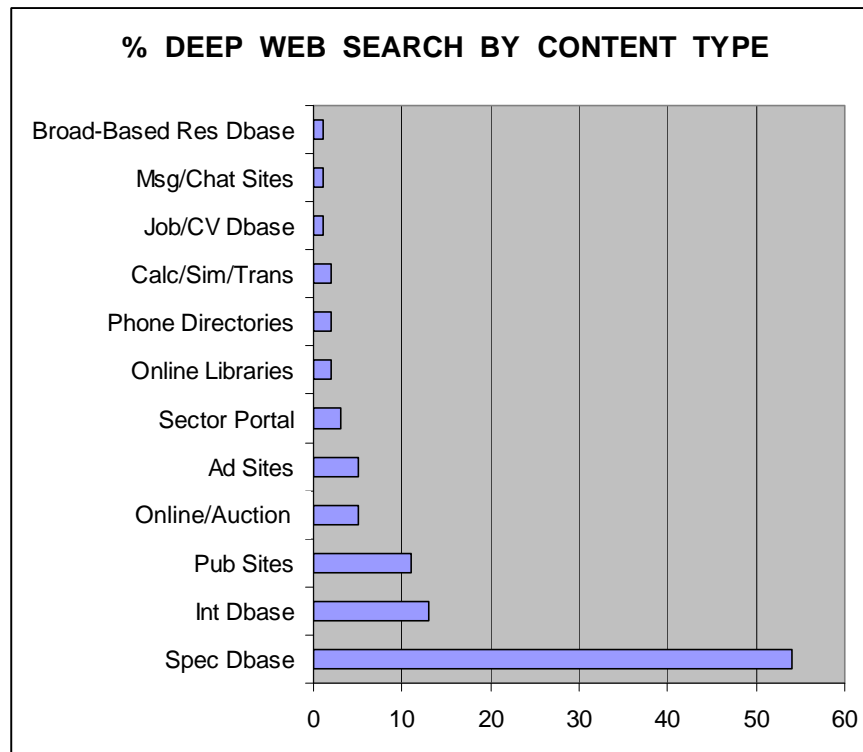
experienced with broad based search engines such as Google, Yahoo, MSN, Ask.com and AOL. These internet search failures add an economic cost to organizations or institutions in terms of loss of productivity, unsuccessful search results, and additional search related salary cost. In an Outsell (2006) survey, it was reported that internet searches of broad based search engines accounted for a 68% success rate. Some 32% of these searches were reported as being unsuccessful.

In a January 2007 survey (source Hitwise), based on searches conducted over a four-week period from a sample of ten million searches, the distribution of searches across broad-based search engines are shown in the following chart.



Olivier Scheffer reported at the 2007 Search Engine Meeting in Boston that “broad based search engines are missing the most valuable part of the Web...” often referred to as the Deep Web or an Invisible Web. What is advocated is “fully customized vertical search that will improve search results.” See discussion below.

Scheffer’s display of deep web search sites by content type shows that some 54% are specialized databases, with 13% internal databases and 11% are publication sites. A full detail is provided below: Source: BrightPlanet



- Specialized Databases 54%
- Internal Databases 13%
- Publications Sites 11%
- Online Sales, Online Auction Sites 5%
- Small Ad Sites 5%
- Sector Portal 3%
- Online Libraries 2%
- Yellow Pages and Phone Directories 2%
- Calculators, Simulators, Translators 2%
- Job and CV Databases 1%
- Messages and Chat Sites 1%
- Broad-Based Research Databases 1%

Source: BrightPlanet

What is vertical search? It is part of a larger sub-grouping of specialized search engines. It is a new tier in the internet search industry that focuses on specific businesses. These search engines attempt to address the information needs of specialized or focused audiences and professions. They target niche audiences. “Vertical search engines contain information in their indexes about a specific topic. They are aimed at people who are interested in a particular area, and deliver to a narrow and much focused audience to the companies that advertise on them,” (Perez 2006). Such engines may be designed for patients, job seekers, travelers, doctors or engineers. Vertical search engines are able to deliver relevant and essential information that is difficult to attain with the use of broad based search engines. Highly specialized vertical search companies may pose the most significant threat to broad base search engines such as Yahoo and Google, but the lack of name recognition and brand awareness makes it difficult to sustain a high traffic flow (Regan 2005). Both Yahoo and Google have established their own vertical search tools to garnish a segment of the vertical search market. The key to the survivability of a true vertical search engines is specialization, such as Answer.com that focuses on specialized research. A key issue is whether the proliferation of vertical search engines can retain customers or whether they will deviate to sites that meet most of their needs. Regan (2005) also noted that “LookSmart.” believes that searching on the web will become vertical and personal as customers search for essential content that may be hobby related or educational in nature. The hope is that information seekers will use the web as they do cable television, favoring specialized channels that address their concerns or interests. This behavior would lead to search engine optimization.

There are uncertainties about the long term impact of vertical searching. Perhaps information seekers may prefer simple search options accessible from a single search system that they trust and are familiar with, as opposed to seeking sites that offer the best access to specific types of information. The information seeker’s level of sophistication, knowledge of the subject, level of education, and the complexity of the question or content sought may all play a role in searching for information. On the other hand, as the internet becomes more populated with information seekers and providers, consumers may search where specific sources of information reside

(vertical search engines) versus a one-stop shop approach where searching may be more convenient and/or easier (broad based search engines) but perhaps less accurate and slower.

Desired Improvements in Searching

Tom Reamy's presentation at the 2006 Search Engine Meeting in Boston addressed "Faceted Navigation" as an alternative to search and browse. Faceted Navigation is defined as the dynamic combination of search and browse. It is intuitive, with multiple perspectives and allows for the processing of compound subjects. There are disadvantages however, such as its difficulty expressing complex relationships and loss of browse context. Faceted navigation allows for more structure, taxonomies, and metadata.

In Mike Moran's presentation to the Search Engine Meeting 2006, he noted that a good search engine should not be the goal; instead, searching should be viewed as a means to an end. He also stated that the search engine goal is not to deliver good results; instead, the goal is to deliver the business value of your web site. Moran supports the view that the information seeker should search the most popular search keywords first, since most search terms are unique. It is the easiest improvement one can achieve. He displays an IBM (2006) table of all queries against unique ones. The results showed that IBM's 1000 most popular queries accounted for only 27 % of all volume. It can be argued that Moran's approach to searching is valid when expert searchers or subject matter experts are conducting the search. This begs the questions as to whether or not a novice searcher would be aware of the popular keywords associated with a given topic or subject. It is very unlikely, that such seekers of information would achieve as good a search result as expert searchers do.

In Andrew McKay's presentation on "The Future of Search Content Synergy" at the Search Engine 2006 meeting, he provided insight into the vast amount of wasted search time. This wasted time may be attributed to the vast amount of electronic data available. The University of California Berkeley, School of Information Management and Systems, estimated that the rapid growth in information is equivalent to 105 billion gigabytes per week, accounting for a 1 % growth per week. McKay reported from an IDC study, that 50% of all online searches are unsuccessful, and he summarized wasted search time as follows:

- 44% of those who conduct a search are not sure what to type in the search engine
- 39% of searchers use misspelled words that account for poor results
- 13% of users do not know what to look for without assistance
- 22% of searches have no result (IDC)
- 5% of searchers navigate multiple pages to seek information (IDC)

Andrew Pace (2007) emphasized the importance of improving or enhancing the information seeker's experience by "making the bibliographic data work harder for the user or by establishing relationships between the bibliographic data and other systems." The author described North Carolina State University's faceted browser interface as an example. The

bibliographic data is decomposed into facets to enhance the search experience. Pace recommended bibliographic data that has the following features:

- A classification scheme or subject thesaurus that enables faceted classification
- A work identifier for books and serials
- Improved name authority for organizations
- Physical description to include weight, height, and width (support remote storage management)

Enhanced gateways are another approach to improve the information seeker's search experience by way of a centralized and simplified search process. There are attempts by institutions and organizations, such as Google, to use their search systems to enhance the information seeker's experience. The enhancements can be achieved by linking bibliographic data from other sources such as WorldCat in an attempt to find a book or document in a local library, or to find the associated bibliographic data about it, which can also be purchased online through a link with a retailer. Also, links have been established with providers such as Google Scholar for books and journals from local libraries where the information seeker can then access the full text online.

Future in Searching

There are fundamentally differing views on the future of searching. Technological advancements in search systems and improvements in the harvesting of information across multiple databases on a global platform will impact the future. DuPuis (2006) suggested two ways to speculate on the future of searching. The first approach is "how we think things are going to be (dystopian)." The second approach is "how we would like things to turn out, (utopian)." The future information seekers 'net generation' and beyond will not have the present level of attachment to journals, conferences, and monographs; but instead, they will have "expectations of simplicity." They have a desire to find rather than search. They are seeking convenience. The author further noted that publishers and database providers are now beginning to accept the fact that information seekers do not care where the information resides as long as they are able to find the information sought. The key is adding value as an information provider.

A discussion of the future of searching must address the fundamental question regarding how to improve information access. This may require interactive and visualization tools by demonstrating relationships among various entities in multi-dimensional forms. How will search engines interface to improve and deliver seamless results? Will human interaction with machine improve, so that the search engines of tomorrow will be able to understand the information seeker's behavior and anticipate the expected search results? Perhaps human-computer interactions and the ability to comprehensively address the ambiguity of images, words and objects will go a long way in enabling unified access to data across multiple platforms. With better understanding of user behavior, improved search results should be realized.

An understanding of the information seeker's behavior in a search setting provides valuable data to search engine and interface developers for the designing of effective and user-friendly search systems. Early studies in the 1970's that addressed information seeker behavior were focused on

the library setting. These studies preceded access to online search tools. As search engine access for researchers became more prevalent, studies were conducted to assess the information seeker's search behavior in an online setting. Bates (1979), studied the ways in which information seekers performed searches. In 1989, the author recommended methods to describe the search process.

Silverstein et. al. (1998) studied the query logs from Alta Vista search engine. They found that the majority of information seekers used very short queries in conducting searches. Spink, Wolfram and Saracevic's (2002) analysis of Excite query logs for the years 1979, 1999 and 2001 revealed that information seeker's search strategies on the web have remained the same, with a few exceptions such as their unwillingness to view more than a page of search results.

Jan Pedersen, Chief Scientist, Yahoo! Search, estimated in his presentation at the 2005 Search Engine Meeting, that there are >400 million internet daily searches generating in excess of \$6 billion in revenue. Approximately 50% of this revenue is associated with the three major players; Google, Yahoo, and MSN.

McKay (2006) summarized the future of searching in the following manner: it will be universal, pervasive and necessary. McKay believes that technical boundaries will disappear and searching will be available "everywhere all the time." It will be universal! Searching will be pervasive! It will, be more proactive than reactive. Finally, searching is necessary, as it will affect all aspects of one's life. Both consumers and information providers (government and business) will have access to more information about individuals. Consumers' demand will increase with greater expectations.

In Rose and Levinson's (2004) study on "understanding user goals in web search," the authors noted that future improvements in web search engines will require a better understanding of the information seekers' behavior, including both how they search and why they search. The knowledge gain would be used to modify the search engines' algorithms and interfaces to improve search results.

Search engines are now building profiles of information seekers which increase revenue through sales advertisements. Peterson (2005) summarized that impact by noting that companies "know what people want to read and the places they want to go. They're getting an unprecedented look at the collective wants and needs of the population..." As the size of the web grows, now estimated at 11.5 billion indexable pages (Gulli and Signorini 2005), it is anticipated that more personal information will be captured (phone numbers, credit card numbers, addresses, purchasing preferences, and products purchased) (Peterson 2005).

Where does searching go from here? The obvious direction is to allow information seekers access to the web anywhere and anytime. The choices for access include cell phones, mobile devices and television. While device access is available, its searching capabilities are quite limited due to bandwidth issues. Perhaps some day television and searching will merge. This would allow viewers simultaneous access to broadcast programs and searching for more information. The current ability to access video within a search is a step in this direction.

Sokullu (2006) believed that internet searching is still in its infancy, as researchers attempt to find better solutions to searching and improved indexing techniques by exploring new horizons. The author identifies three trend areas in the search industry: user interface (UI) enhancements; technology enhancements; and approach enhancements (Vertical Engines).

Bourdoncle (2007), in his discussion of user interface issues and challenges, suggested that consumer user interfaces are too simplistic (search/browse result list/next page). They are good for unstructured web pages. The author believes that what is needed is a unified user experience to support the many incompatible products. There is a need for a universal browsing tool for semi-structured information.

Mostafa (2005) believed that online search engines are poised for major enhancements that will change how we find what we need. The results that are achieved now are partly due to the deeper searching that is occurring as new search engines are able to refine their processing of increasing volumes of data available on the web. Mostafa noted that search engines such as Google have mastered two major hurdles in information retrieval; that is, “the ability to handle large scale web crawling tasks, and indexing and weighting methods have produced superior ranking results.”

Future searching will go beyond the conventional computing platforms. Search capabilities will be embedded in entertainment equipment such as game stations, televisions, and high-end stereo systems. The author anticipates search technologies as playing a major role via “intelligent web services” in such activities as driving a car, designing a product or even in the way one will be listening to music. These changes will create a new market for new business deals that will result in an expansion of online published materials such as video, audio, and text. “The next generation search technologies will automatically include more powerful tools, combining search functions with data mining operations, which will be able to look for trends or anomalies in databases without actually knowing the meaning of the data. Advances in data-mining and user interface technologies will allow a single search system to provide a continuum of sophisticated search services that are integrated seamlessly with interactive visual functions. The application of the advances in machine learning and classification techniques will result in improvements in the categorization of web content. The net result will be easy to use visual mining functions that will add a highly visible and interactive dimension to searching. The information seeker will be able to search through multiple data repositories by using visually rich interfaces that focus on broad patterns in information rather than picking out individual records, Mostafa 2005.

Role of Catalogers and Indexers

The debate over the role of catalogers and indexers is not new. With increasing technological improvements, the debate intensifies. Technological improvements have led to an increase in full-text and retrieval search systems available for the conduct of inquiries. For the past two decades, there has been a debate over the future role of human indexers. Increasing cost for the labor intensive effort of indexers makes full text retrieval a more attractive option, (Blair and Maron, 1985). There is also the argument that indexers are often both inconsistent and

ineffective. Don Swanson's pioneering study in 1960, evaluated the feasibility of full text search and retrieval. He concluded that "text searching by computer was far more effective than conventional retrieval using human subject indexing." These views were also supported some ten years later by Salton (1970). Researchers and information providers are still debating the issue four decades later. A new dimension to the debate is the economic cost associated with the labor intensive effort.

Over the past decade, there has been a proliferation of digitized data available on the internet in full-text search systems. Information that was only accessible in public and corporate libraries, institutions of learning, federal, state and local government offices, can now be retrieved by both novices and expert searches from the convenience of their homes and offices. The shift in access is partly due to the vast array of data now available on free web search engines. An added benefit is their ease of use. Novice searchers are now relying less on their public librarians for support in finding answers to questions mostly because they can now search the web at their convenience without leaving their homes.

The shift to a digital era begs the question as to what the future role of catalogers will be. Institutions have begun to evaluate the traditional role of library cataloging. The great detail and expenditure to perform descriptive cataloging must be weighed against the economic benefit to organizations in continuing down this path. In a 2006 speech by Deanna Marcum, Associate Librarian, Library of Congress, she noted that the institution spends \$44 million per year on cataloging functions. The author questions whether the institution should continue down this path in light of 'digital information, internet access, and electronic key word searching.' As more information seekers rely on Google, Yahoo, MSN and other internet search services, library catalog usage and value will perhaps decline. Where do we go from here? Marcum asked a question. "Do we need to provide detailed cataloging information for digitized materials, or can Google be viewed as a catalog?" There are certainly large volumes of data now available on the web (both scholarly and non-scholarly, full text documents and bibliographic data) that reduces the information seeker reliance on library catalogs for discovery. This debate is further complicated by the Google declaration some two years ago of its' intent to create a global virtual library by organizing the worlds information. Efforts have been under way with agreements with several institutions of learning and the New York Public Library to digitize selected works from their collections that would be made available to information seekers worldwide, through Google. The fundamental question is, what do information seekers need from online catalogs in the twenty first century? The high cost of cataloging and its shrinking use may dictate its future.

In Calhoun's (2006) study on the changing nature of cataloging she suggested that there are "prevailing strategies for integrating the catalog with other discovery tools, many research libraries leaders, staff members and university faculty members are not ready to accept this change." The author refers to initiatives such as Google Book Search, Open WorldCat, and RedLightGreen as promising in exposing research libraries collections on the web. There is some doubt as to its attractiveness to scholars and students. Search engines have become the primary sources for scholars and students to begin their inquiries. Calhoun suggested that research library catalogs reflect only a small portion of the ever expanding universe of scholarly

information. This would therefore decrease the demand for catalog usage among these two groups.

The maintenance cost to support catalog records is huge. In 2005, ARL libraries spent an estimated 239 million dollars in labor cost for technical services support. Regardless of the cost, Calhoun believes that these records will play a major role in discovery and retrieval for sometime to come. In 2005, OCLC estimated that there were some 32 million research library books to be digitized.

Interviewers in the Calhoun study identified several unique advantages of catalog usage to information seekers. They are:

- Allows for bibliographic control
- Contains good metadata to describe and collocate related items
- Supports browsing
- Offers predictable and consistent structure of catalog records
- Provides detail information about items and their status
- Manages large collections
- Supports delivery of those collections to users
- Provides access to information not available on search engines

Calhoun (2006) suggested that catalogs of the future will be a “link in a chain of services that enable information seekers to find, select, and obtain the information objects they want. Future catalogs will be required to ingest and disperse data from, and to, many systems inside and outside the library.”

There is a down side to information seekers when they by-pass the online catalog for broad-based search engines which offer ease of use (such as Yahoo, MSN or Google), in their search for answers. Bates (2003), warns that information seekers will use information even when they know it to be of poor quality or unreliable, so long as it is easy to find. The key is ease of use and access to information. Byrum (2005) suggested that library catalogs need to provide access to more content with enhanced interfaces to attract information seekers. The catalog is limited in scope, with emphasis on print. This is a drawback to its use, (Medeiros, 1999). The commonly held view is that online catalogs are hard to use due to their outdated interfaces.

In Thomas Mann’s (2006) critical review of Calhoun’s report of the changing nature of cataloging, he pointed out that there is a clear difference in the needs of scholars and those of “quick information seekers.” Listed below are the points which Mann has used to support the need and importance of cataloging which he views as a valuable tool for scholars.

- They seek clear and extensive overview of all relevant sources.
- They are concerned that important, significant sources not be overlooked.
- They prefer to avoid duplication of prior research.
- They are interested in cross-disciplinary connections to their work.

- They wish to find current books on a subject categorized with prior books on the same subject.
- They prefer mechanisms that allow the recognition of highly relevant sources.
- They would rather avoid having to sort through huge lists or displays.

In Markey and Burke (2007) comments from the “working group meeting on the future of bibliographic control,” the authors stated that “information seekers need additional rich data other than the bibliographic catalog to find information. Multiple access tools for information discovery are also needed. These tools include general search engines that use keywords as the access methodology to more specialized systems such as faceted browser interfaces. Bibliographic data should expand beyond English language searches and structures.” Markey also reported that information seekers’ use of bibliographic data (online catalog) is affected by their system knowledge, domain expertise, and their procedural knowledge. The author noted that 77% of users have low system knowledge and low domain expertise/procedural knowledge. This group is defined as “double novices.” At the other end of the spectrum, only 5% of information seekers demonstrated high system knowledge and high domain expertise/procedural knowledge. They are defined as “double experts.” Markey recommended that enhancements to retrieval systems and bibliographic catalogs should be focused on helping the “double novices.” Markey and Burke suggested that a “double expert” someone who has specialized knowledge of a discipline, could be a “double novice” once that person attempts to conduct a search outside of his area of specialization.

DATA ANALYSIS

The search methodology study participants were grouped into **five sub-groups**. Each sub group interview responses were divided into **seven major categories**, which are displayed in **35 tables**, see appendix: A.

The sub groups are:

- CENDI Member Agencies (An Inter-agency group of Senior Level {STI} Executives and Managers from Federal Agencies)
- DOD Organizations and DOD Contractors (Library Professionals)
- University Information Science and Computer Science Professors
- Information Science Organizations
- Other Libraries

There were 48 participants from 29 organizations and agencies. Participants included: Information Science Professionals (senior managers, technical information specialists, and librarians) from the Scientific and Technical Information (STI) community within the federal government, Reference Librarians and other information providers from the university community, University Professors from information science and computer science departments from several universities, Professionals from various information science organizations and companies, and Information Professionals from non-CENDI Federal agencies and Government supported organizations.

For the purpose of discussion and analysis, the responses from each group are displayed in table form.

The seven major categories are:

- Preferred Method of Searching
- Searching Methodology...Full-Text, Metadata, Other
- Limitations in Full-Text and Metadata Searching
- Search Systems Performance...Measures
- Improvements in Retrieval Effectiveness
- Future Role of Catalogers and Indexers
- Improving Search Results...Role of Metadata and Full-Text

PREFERRED METHOD OF SEARCHING

Table # 01: Responses from the 20 CENDI participants to the questions relating to their preferred method of searching and the reason for their choices. Respondent's views were mixed regarding their preferred method of searching.

Question # 17: What is your preferred method in searching databases for access to government information?

_____ Full Text _____ Metadata _____ Other _____ No Preference
_____ Specify

Question # 18: Explain the reason for your choice?

Summary Responses:

Participants acknowledged the benefits derived from each method of study. The method used varied with their knowledge of the subject, the richness of the database, how well a database is indexed, participants knowledge of the information being searched, the comprehensiveness of a database, and its ease of use.

In some instances, participants preferred to do an initial full-text keyword searching in their first attempt in finding information. This approach provides an initial survey of the number of hits that can be derived. Follow-up searching may take the form of metadata searching with more narrow and precise words, terms or phrases.

One participant noted that the method of searching used will depend on the type of information being sought. When looking for a specific fact, full text searching may be the only way to find it; on the other hand, when looking for specific known documents, author or title, metadata is used. The participant leans more towards metadata searching for search engines that have that capability.

Participants' responses also included the following: when data is inputted correctly in a metadata search system results are more relevant; also, a combination of both search systems is advocated. A participant recommended the combination of Boolean fielded searching including controlled vocabulary, Boolean full-text searching, and algorithmic full-text searching

Another participant in support of full-text searching noted, that the search terms that is used may not be part of the controlled vocabulary, and the results may be minimized or fruitless. He further noted that metadata is not cost effective, and that results may reflect poor cataloging. The belief is held that recall for full text searching will always be better than that of metadata searching. Full-text searching with metadata tags is advocated.

It was also stated that by using a “wide net approach” to cast to see what type of information is available perhaps through full text searching of Google, such as Google Scholar, sets the stage for more precise searching that may include metadata searching.

Summary Responses:

Table # 02: Responses from 14 DOD Organizations and DOD Contractors participants.

The table provides participants’ responses to the questions relating to their preferred method of searching and the reason for their choice. Respondents’ views were also mixed. Participants’ acknowledged the benefits derived from each method of study. Responses varied from those who preferred full text searching because of its ease of use, to those who acknowledged the added benefit when both methods of searching are used. One major draw back noted in using full text searching is the large number of hits derived.

Those who favored using full-text searching when accessing government information noted the ease of use. Also stated was the view that “good metadata” is not widely supplied with government information and searching by metadata requires you to know the appropriate government jargon to match. Another approach is to conduct a full text keyword search, and if a relevant article is found, then the use of subject field is applied to find more relevant information.

One participant noted that full-text searching capability is immensely helpful when looking for a needle in a haystack - when the classification or structure or hierarchy is not known but a small amount of very precise information is available

An advocate of full-text searching pointed out that “good metadata” is not widely supplied with government information and searching by metadata requires you to know the appropriate government jargon to match. Another view held is that of using full-text searching as a vetting process, i.e., Google Scholar; this is followed by application of metadata searching to improve search results.

Summary Responses:

Table # 03: Responses from six University Professors participants.

The table provides participants’ responses to the questions relating to their preferred method of searching and the reason for their choice. Most respondents preferred having a choice in using either full text or metadata searching when accessing government information. One participant believed metadata searching was more suited for accessing government information, due to the complexity of the information. Another, participant thought that the method of searching will depend on what type of information is sought.

Summary Responses:

Table # 04: Responses from six Information Science Organizations participants.

The table provides participants' responses to the questions relating to their preferred method of searching and the reason for their choice. Participants' responses included the following: the combination of both search methods gives the best of both worlds and may support both high precision and high recall requirements, full text is easier and faster when ones' knowledge of the system is limited, the more one understands the system the more effective metadata can be, and ones' preferred method depends on the nature of information required. A draw back noted when using metadata searching is the lag time it takes for new terms to be included in a controlled vocabulary.

One participant believed that the preferred methodology should depend upon the nature of the information required at any particular time. There may be times when full-text is absolutely required and other times when an amplified "abstract" or surrogate of the full text (ie. one that contains an intimation of the conclusions reached in the research paper or a graphic that illustrates a particular region) will be adequate to the purpose.

Another participant believes that if you don't know the system then the easiest, fastest way is full text. "The more one is a power user and understands the system the more effective metadata can become."

Summary Responses:

Table # 05: Responses from two Other Libraries participants.

The table provides the responses from the two participants' to the questions relating to their preferred method of searching and the reason for their choice.

One participant believed that more precise search results can be obtained when using metadata searching. The other participant used both methods to search. The searching method was determined by what information was sought.

SEARCHING METHODOLOGY...FULL-TEXT, METADATA, OTHER

Table # 06: Responses from the 20 CENDI member participants. The table provides participants' responses to the questions relating to the status of searching methodology.

Statement # 01: Scholars often refer to full text searching as searching devoid of controlled vocabulary, taxonomies, subject classification, metadata, etc., when in fact; most full-text databases often incorporate some form of classification, structure, complex search algorithms, bibliographic fields and abstracts.

Please Comment.

Summary Responses:

- Support the view that there is a mix which varies between databases
- XML is dominating the landscape as the markup language of choice
- It is what goes on behind the scenes in the technology of building the database that is dominating and is changing the input requirements for putting the data/information in the database
- Better searching is enabled by richer databases, however, bringing the collection under bibliographic control which would further improve search capabilities is unaffordable
- A good XML structure adds value to the results
- Clearer and better information about the applications would improve searchers' understanding
- Having an access system that can accommodate both full text and metadata searching is important for many organizations
- Do not agree that a full-text search takes advantage of classification fields, abstracts, etc.
- There are a number of technologies that use various techniques to improve searching.
- They use complex algorithms and formulations, for example FAST. This search engine looks at word relationships but Google does not ...maybe three fields
- In the perfect world, controlled vocabulary would be universally applied and would provide optimum search experience
- Taxonomies can now be controlled, or system generated. Either way, they can be used to facilitate full text searching
- Full-text databases don't always incorporate legitimate structure, algorithms, etc.
- Maybe some do, but others, for personal or economic reasons, have developed algorithms that are almost considered trade secrets
- True Statement!
- True! Most full-text databases are no longer pure full-text devoid of structure but actually have metadata searching features

Statement # 02: Early web search engines relied on Boolean methodology in meeting full-text searching needs. Later web programmers gradually began applying metadata, taxonomies, and algorithms. We now find bibliographic and full-text information combined. Is the real issue therefore, what recipe of metadata, taxonomies, and algorithm to apply.
Please Comment.

Summary Responses:

- The real issue at this time, and given current technologies, is optimizing the recipe or mix of metadata, etc.
- That recipe also depends on the data/information types that comprise the database
- Take advantage of whatever recipe that the databases being searched will support
- Applying commonly understood and interoperable indexing aides is the very essence of being able to increase the value of web search results
- It is preferable to have a rich mixture of metadata and taxonomies that have cross-walks between them
- The issue is not which recipe to apply, but rather how to present search choices most simply
- In database environments like bibliographic catalogs, full-text journal databases, and other “deep Web” databases not searchable via web search engines, a combination of full-text and metadata searching is prevalent
- The real issue may just be the amount of data/volume of data/information we have to deal with from a user/retrieval point of view.
- Yes, by using algorithms and some metadata
- Note that the bibliographic databases started with metadata and only later turned to full text, mostly limited by technology at the time
- By mixing full text and meta tagging, search results can be improved, but the taxonomy must be consistent and consistently applied
- Taxonomies can now be controlled, or system generated
- That could be the case that we are approaching more of a blur in searching methodology
- We are now combining searching methodologies.
- An elusive special recipe of metadata, taxonomies, and algorithms is not going to generate 99% accuracy for searchers

Summary Responses:

Table # 07: Responses from 14 DOD Organizations and DOD Contractors participants.

The table provides participants’ responses to the questions relating to the status of searching methodology

Statement # 01 Responses:

- As search engines evolve, there is much better control and more appropriate retrieval generated.
- With the combination of bibliographic and full-text data, we can achieve increasingly better search results.
- Some sort of algorithm must exist for full text searching to accomplish its tasks for the searcher.
- They do have controlled vocabulary hidden.
- I believe this to be true.
- Search algorithm as applied to full text searching is very different from the kind of hierarchical, taxonomic, classification-based approach one takes when reviewing the literature for a specific topic.
- Totally devoid, but I do like to use a controlled vocabulary. On topics that one is not knowledgeable in, by consulting a controlled vocabulary this becomes a valuable tool.
- I don't believe that full text databases incorporate classification or structure.
- Most full text databases do not have a good controlled vocabulary. Do have metadata, but not necessarily controlled vocabulary.
- Yes, some form, such as limiting to a certain field, which helps.
- Most search engines do look and weigh such fields as title, if they are supplied as a metadata tag, and others can be added to the calculation of relevance as appropriate
- Few organizations exploit both full text and metadata searching capabilities. In the legal, genetics, technical fields, it is important to have both searching approaches.

Statement # 02 Responses:

- The combination of metadata, taxonomies and algorithm may vary in terms of the subject matter to be searched.
- Yes, the real issue is what mix of metadata, taxonomies, and algorithms to apply.
- If both the full text and metadata are used for retrieval, then I think there needs to be some method of limiting search results; i.e. by author, data, or title.
- Agree with identifying the recipe of metadata to apply.
- There were sophisticated search engines before there was a web. DTIC, Dialog, and many others pioneered in Boolean search.
- Yes, with the goal of keeping the widest range of approaches available to provide flexibility for the user.
- Agree. I find value in the taxonomies.
- It is important to have an underlying structure with taxonomy and algorithms.
- There is no "one size fits all" solution. In some environments, the use of metadata and taxonomies may be appropriate; in others, such a fixed structure is not

appropriate because of the time and effort required to establish, evolve, and maintain taxonomies

- Taxonomic structures tend to be frozen in time and thus antithetical to discovery; they tend to be one person or group's view of the information's organization; and they are generally implemented with little understanding of the end users' information seeking behaviors.
- Searches need subject indexing. With better and more complex algorithms, though expensive, one will get better data extraction.

Summary Responses:

Table # 08: Responses from six University Professors participants.

The table provides participants' responses to the questions relating to the status of searching methodology

Statement # 01 Responses:

- You offer a premise that you purport is a fact (most full-text databases incorporate...) but whose truth I am not at all convinced
- Even if your premise holds, the most you could conclude is that: either most searches ignore available information in the databases or that most searches of those databases are not, by definition, full text searches
- The real issue is what algorithms to apply to a given corpus for a given user community
- The better approach is to know the community that will be searching the collection, know how to build an interface to let that community specify what they are looking for, and then build in algorithms that fill in search limiters previously found useful.
- The interface is the issue. The simpler the interface, the less knowledgeable the user, the more work has to be done by the search algorithms.
- The users' take for granted that the search system will take care of the problems to make the system work.
- While a full text database may incorporate some form of classification, its search engine may not always allow searching that way
- Page rank or probability ranks are better than controlled vocabulary and metadata

Statement # 02 Responses:

- I agree partially. I still sometime prefer to use Boolean methodology in search (like use of phrase search in Google)
- Again, the premise seems at best marginally related to the conclusion
- Automatic metadata generation and indexing is the way to go. Using taxonomies and metadata can be viewed as a way to reduce the noise in search

- Of course, as the search engines say---it is the secret sauce that differentiates search services

Summary Responses:

Table # 09: Responses from six Information Science Organizations participants.

The table provides responses to the questions relating to the status of searching methodology

Statement # 01 Responses:

- The point that new technologies are not void of knowledge structure is correct
- The better the structure applied to data the more likely a search will turn up relevant material
- It depends on the search engine. Google still does not use metadata to the extent that Yahoo does
- It wasn't until Google emerged that we started seeing the massive Web audience introduced to the idea of special algorithms as a part of the search environment
- This is so true when you can put in your search elements that make use of the metadata such as domain name
- While I agree with the statement, the problem is the information itself and the way it is displayed

Statement # 02 Responses:

- Yes, it is recipe of the mix but there is also need for and existence of continuing advances in the underlying models on how to apply them
- Yes, of course, it is the mix of all these that are applied. The right mix isn't easy to achieve. In addition, it also depends on how well the metadata, taxonomies and algorithms meet the needs of an increasingly more diverse audience
- Yes, with the understanding that there will be an on-going challenge for content providers to develop different recipes according to the needs of a specific community of practice

Summary Responses:

Table # 10: Responses from two Other Libraries participants.

The table provides responses to the questions relating to the status of searching methodology

Statement # 01 Responses:

- Yes, some form, such as limiting to a certain field, which helps.

- That's true in part, but unless users are aware of the controlled vocabulary terms used in the full-text database, they still are whistling in the dark.

Statement # 02 Responses:

- A recipe that limits your results to something meaningful is what we want.
- Agreed! The problem is that people, including library administrators, want everything to work like Google—plug in terms and supposedly you're set. The work it takes to set up taxonomies and provide metadata tags is pretty staggering, especially if you are trying to do it retrospectively.

LIMITATIONS IN FULL-TEXT AND METADATA SEARCHING

Summary Responses:

Table # 11: Responses from 20 CENDI Agencies Participants.

The table provides responses to the questions relating to limitations in full-text and metadata searching

Question # 22. What are some of the limitations in using full-text searching?

- Results can be overwhelming, devoid of context, less relevant, and not very time efficient
- Relevancy is often a problem with full-text searching
- Its relative imprecision compared to retrieval based on a controlled vocabulary indexed system
- End users' inexperience with full-text search strategies, such as the need to include variant forms and synonyms of a keyword, might lead them to feel dissatisfied with their search results
- Improper classification of documents, slow response time, difficulty in presenting customized results to users
- Few drawbacks! It can give you more information than you desire.
- Not finding what you are looking for because of too many hits
- Lack of synonyms. Not being able to differentiate.
- Specificity is lacking
- Large number of hits, or false hits.
- Lots of bad results with relevancy that is not meaningful
- One issue is that searchers don't optimize their search strategy
- False results.
- Words searched or retrieved may not be relevant terms. For example; military terms or acronyms
- Relevance. Too many hits.
- Typically receive many non-relevant documents

- If it is an algorithmic search of a full-text database, the disadvantage is lack of precision and control
- Full-text searching is completely at the mercy of the author (or scanning software) and errors that they made
- Using the wrong word(s)
- Not knowing the right word(s)
- Irrelevancies, too much stuff, etc.
- Lack of precision ... unless the words you are using are really precise themselves
- High recall, low precision
- May need to look at a lot of records before you find the relevant one
- Normally full text searching does not yield many relevant results
- Lead's to large irrelevant results

Question # 23. What are some of the limitations in using metadata searching?

- You are totally dependent (and at the mercy) of whoever created the metadata
- Other than for the classes of documents or information I mentioned above I see little use for metadata in today's world
- The main drawback using metadata searching is that few can afford to create the metadata
- It is useless if the user does not know the structure and meaning of the metadata
- Lack of controlled vocabularies by the author
- Results may not be as rich as with full-text included
- Metadata is also expensive to create, assign and maintain, so its quality varies greatly from database to database
- End users' inexperience with metadata searching
- Less prominent topics that appear as part of the document or data but are excluded from the metadata
- The end user often has to be a more experienced searcher to do effective metadata searches
- To improve results, and typically a user has to understand the scope/intent/content of the repository better than with full-text searching
- No drawbacks
- Terms one is looking for may not appear in the citation
- You can only use subject terms from the title, abstract, or controlled vocabulary
- Inconsistency in the controlled vocabulary
- Terms change over time
- May not capture all results
- To create the metadata is expensive
- User's unfamiliarity with metadata rules or lack of understanding may result in poor output
- Terms may not be imputed correctly or consistently
- Indexers may not be picking up the best terms for the documents

- Misspelling when inputting data
- One may miss the most important document, since one is relying on the work of the cataloger and indexer
- Metadata databases do not allow you to find quotes nor every mention of a word or phrase in the full-text
- Not knowing the vocabulary or understanding the concept
- Requires more education and thought
- The taxonomy/thesaurus/metadata schema needs to be accessible to users or they won't be aware of them
- These tools also need to be pretty sophisticated (lots of references) or they won't pull up arcane terms
- One relies upon the expertise of the human inputting the metadata
- Not always clear how system works.
- The way we describe terms.

Summary Responses:

Table # 12: Responses from 14 DOD Organizations and DOD Contractors participants.

The table provides responses to the questions relating to limitations in full text and metadata searching

Question # 22. What are some of the limitations in using full-text searching?

- As a single approach, it may not draw together the elements that will most quickly pinpoint a document
- Natural language idiosyncrasies, use of slang and jargon, abbreviations or acronyms that can have multiple meanings, misspellings
- Getting too many hits because of citation listings
- Terms may only appear in the title which then results in a low relevancy
- You will retrieve irrelevant results
- Increase the chance of getting spurious results. Time consuming
- Search terms may not match jargon or business-specific terminology
- Content may include a number of synonymous terms, depending on the author, where uniform use of terms would be better
- Users may use a variety of ways to search for the same content
- Mismatch of user terminology with jargon used in the content or with something that needs a fairly exact match, such as a form number
- Acronyms used in the content may not be familiar to users
- Poor precession and recall.
- Issues of awareness, performance and usability.
- Having to "or" every way a word is used in order to get good results.
- High recall. Irrelevant material.

- Large number of hits.
- Increases the chances of getting spurious results
- Time consuming

Question # 23. What are some of the limitations in using metadata searching?

- A relative little used term may not be included in metadata string if the metadata creator did not choose to include the term.
- Controlled vocabulary, unfamiliar with thesaurus
- Not being able to find documents on specific subtopic; i.e. M28 projectile info not found in a metadata search that a document under projectiles.
- Whether the end user has the ability to relate to the subject terms used.
- The user ability to understanding the concepts in a metadata search
- Lack of precision and flexibility is a possibility
- Inconsistency, expensive, time consuming, difficulty in keeping terms up to date in certain disciplines due to constant changes.
- Controlled vocabulary is slow in updating.
- It takes a while for new terms to be accepted.
- Also, the use of author key words.
- It may eliminate relevant data.
- Taxonomy may not be created well.
- It must support system for which it was developed

Summary Responses:

Table # 13: Responses from the six University Professors Participants.

The table provides responses to the questions relating to limitations in full text and metadata searching

Question # 22. What are some of the drawbacks in using full-text searching?

- Too many hits, term ambiguity
- Low precision, low speed, unfriendly interfaces
- Too many hits, term ambiguity

Question # 23. What are some of the drawbacks in using metadata searching?

- Not user friendly
- Low recall & precision, unfriendly interfaces, cost of acquiring accurate metadata
- Too much information noise while possibly missing out important information
- When everybody becomes information literate, metadata searching is not a problem

- Too few hits, missing categories, term ambiguity

Summary Responses:

Table # 14: Responses from the six Information Science Organizations Participants

The table provides responses to the questions relating to limitations in full text and metadata searching

Question # 22. What are some of the limitations in using full-text searching?

- Getting results that have nothing to do with the user's thoughts in the search query but are in fact accurate in the use of the terms used in the query
- Precision of search -- Fine tuning the search well enough to get what is really wanted
- Pure full-text searching is very dependent on the search engine
- Language is the biggest drawback and consequently the volume of content retrieved
- Speed and use of system
- The information layout. Researcher has to scan the full-text to find the needed information

Question # 23. What are some of the limitations in using metadata searching?

- Expense of applying the metadata and allowing only the term deemed the preferred term in the search itself
- Human error in the construction of metadata
- Metadata searching can sometimes be too precise
- Limited understanding on the part of the user as to what fields are included
- Inconsistent data
- May limit the task at hand
- Researcher must understand the way the information is presented

Summary Responses:

Table # 15: Responses from Two Other Libraries Participants.

The table provides responses to the questions relating to limitations in full text and metadata searching

Question # 22. What are some of the limitations in using full-text searching?

- Lack of precision
- Lack of precision ... unless the words you are using are really precise themselves

- These kinds of searches pull up tons of false hits

Question # 23. What are some of the limitations in using metadata searching?

- The taxonomy/thesaurus/metadata schema needs to be accessible to users or they won't be aware of them
- These tools also need to be pretty sophisticated (lots of references) or they won't pull up arcane terms.

SEARCH SYSTEMS PERFORMANCE AND MEASUREMENT

Summary Responses:

Table # 16: Responses from the 20 CENDI Member Participants.

The table provides responses to the statements relating to search systems performance and measurement

Question # 09: Scholars often comment that if searchers had access to more accurate search systems, they would be more successful in their search results. Could it be that search systems are already “good enough,” so that a more accurate system would provide at best only marginal improvements?

Please Comment.

- I think this is often true.
- Accurate searching implies being inside the searcher's head. Only the searcher knows what he or she wants and sometimes they don't even know, which, is the discovery part of what we do.
- Emerging generations of search systems will provide enormous benefits
- Just as a reference librarian can aide even the most experienced researcher, the refined nature of improved search will certainly assist in getting searchers to the right result
- I disagree. I think that search engines are better than what they were, but more improvement is obviously needed, particularly for granular levels of content
- If “more accurate” can be interpreted as “more comprehensive,” I agree
- I think that improving search systems will continue to benefit power users
- Perhaps what is needed is not more accurate search systems, but rather improved search tips and help to guide the searchers in conducting more effective searches
- Results are not accurate or users don't retrieve what they desired
- There are significant improvements in search results by improving the algorithms and by exploring more data
- Search interfaces need to be designed to allow different kinds of searches – retrieval of a specific document, all on a topic, a few good articles, specific fact
- Search systems are getting better but they are not good

- It's more than an accurate system. There are other important issues such as ease of use, recall, precision, intuitive interface, etc.
- I don't think that we will ever believe that our search systems are good enough! As our systems become more advanced, our expectations as users become higher
- Search systems can always be improved. Users don't care about the search system as much as they do about quality of the interface
- You can improve the user interface to help the user to easily create better search statements
- You might also improve the catalogers' application of the controlled vocabulary by giving them more time or training. Pure full-text databases, on the other hand, can try to improve their accuracy only by modifying their fuzzy algorithm. Once they incorporate metadata they become a hybrid with more options
- Search engines are always making improvements to their algorithms.
- More metadata tagging, means better results.
- There are significant improvements in search results by improving the algorithms and by exploring more data.
- Full text search engines need to go back to metadata to get more specific data, such as using author searching.
- It depends on what one is looking for.
- A user might well want to emphasize recall rather than precision.
- Search interfaces need to be designed to allow different kinds of searches – retrieval of a specific document, all on a topic, a few good articles, specific fact
- Search engines need to be flexible to allow different interfaces and capabilities for different needs
- Search systems are getting better but they are not good
- There is not one universal and valid relevancy ranking method.
- It's more than an accurate system. There are other important issues such as ease of use, recall, precision, intuitive interface, etc.
- I don't think that we will ever believe that our search systems are good enough! As our systems become more advanced, our expectations as users become higher.
- Search systems can always be improved! Users don't care about the search system as much as they do about quality of the interface.
- Metadata databases tend to do, are supposed to do, exactly what you tell them with 100% accuracy. Pure full-text databases, on the other hand, can try to improve their accuracy only by modifying their fuzzy algorithm. Once they incorporate metadata they become a hybrid with more options.
- Searchers could yield more target results. For some searchers, the more target the search is the happier they are.
- I would like to see improvements on interface design and usability, making the search system more seamless.
- Sure. Some of them are good enough and some are completely inadequate. I think they all really need evaluation on a case-by-case basis. The characteristics of the evaluators have to be documented as well.

Question # 10: What are some of the fundamental flaws in measuring a search engine performance, and how does one overcome these issues?

Please Comment.

- There is considerable improvement possible with ease of use, and ability for the user to customize the search environment and results
- I am interested in improving search engine performance. Major further improvements are coming.
- Build an autonomous, intelligent agent that learns from both user actions and from the information content of queries and documents
- If measures of volume and speed of retrieval, precision and recall, as well as usability testing can be supplemented with more human-intense follow up, then performance can be tested more fully.
- Providing user education and virtual support to less experienced users to improve their search results...
- Search engine performance is ultimately measured by what a user expects the search query to return. This is flawed since we do not think the same, expect the same results, and/or have different cultural/educational backgrounds.
- It is hard to get test data sets of significant size in order to determine relevancy
- Not sure how to measure! Precision and recall aren't perfect. You do not know how to measure until you know all the hits that match the intent of the query
- Search engines are judged by their speed or results. This does not mean right results.
- There is no simple way to measure the quality of the result
- User evaluations are probably the most important measure
- Some Web publishers purposely use incorrect metadata so that their information will be retrieved by searchers. We cannot overcome all of these issues since many search systems are motivated by economics
- Search engines tend to measure only their hits. You don't know if the user got hits! You only know that they got results.
- Probably the biggest possible flaw in measuring search engine performance is using searcher satisfaction as a measure. Searcher satisfaction is a good measure of a search interface, not of retrieval
- It is hard to get test data sets of significant size in order to determine relevancy. One needs a large data test to get good results.
- There is also different rating and ranking among different search engines!
- Not sure how to measure. Precision and recall aren't perfect. There is no way to get perfect precision or perfect recall, though you could get perfect retrieval if you just retrieve the whole collection.
- Search engines are judged by their speed or results. This does not mean right results. It is more a question of, do you find what you need.
- User evaluations are probably the most important measure. Traditional measures are recall and precision.

- One could question the purpose and the accuracy of the company or person who sets the algorithms for a given system. Also, one may question what factors are used in determining the relevancy ranking.
- Search engines tend to measure only their hits. You don't know if the user got hits! You only know that they got results. That is all the user status provides.
- Probably the biggest possible flaw in measuring search engine performance is using searcher satisfaction as a measure. Searcher satisfaction is a good measure of a search interface, not of retrieval.
- Too many results from search! Presentation of results! Added value! Searchers can make own judgment from results, eg, Google.
- Searching only a selection of material (web search engines generally search only top level), relevancy ranking, minimal controlled vocabulary/ indexing (esp. in database searching), lack of multimedia searching within a document.
- Not sure; the wrong evaluators? Targeting the evaluation to the proper user group.

Summary Responses:

Table # 17: Responses from 14 DOD Organizations and DOD Contractors participants.

The table provides responses to the statements relating to search systems performance and measurement

Question # 09: Scholars often comment that if searchers had access to more accurate search systems, they would be more successful in their search results. Could it be that search systems are already “good enough,” so that a more accurate system would provide at best only marginal improvements?

Please Comment.

- In scholarly research “good enough” is not good enough
- When a user or any searcher better understands the system that they are using, the better they can achieve results they expect
- I think this is true. I can't foresee any improvement to a full-text/metadata search that would generate better results
- There is more room for improvement.
- They probably are good enough – it's just that there are so many of them. The days of one overarching databank – such as Dialog – serving as an exhaustive federated search tool – are gone
- Again, I should think it would depend on the topic or area to be searched
- Potentially but there is always room for improvement and perhaps solving the concern of locating documents that are assigned low relevancy ranking
- There is clearly always room for improvement in search algorithms but the ideal system is basically impossible because of the fact that a typical search (which is less than two words) can often be interpreted by humans in a multitude of ways. Much of the solution to search problems comes down to understanding

information seeking behaviors and providing content to guide the user in the discovery process.

- Search systems are not good enough. Yes systems need to get better with more relevant results. Both accuracy and usability need to be improved. Google provides searchers with sociability with their search experience.
- Librarians know what they are doing. Search interfaces are only marginal! We need advanced search button.
- Don't know. Too much recall in full text searching. For example, INSPEC (Electrical Engineer, Computer Science) database indexed in many ways. It helps in post processing; some databases have begun to do so.
- No! Improve interface and user interaction. This will improve search results.
- Again, I should think it would depend on the topic or area to be searched. If I am looking for test results for the effect of VX on polycarbonate materials at low temperatures, for instance, I would benefit from the most accurate system available. If I am trying to survey or identify technologies used in stand-off detection, I would not want to limit those results unnecessarily - I'd want a very inclusive search and would therefore NOT benefit from exquisitely precise searching.
- Potentially but there is always room for improvement and perhaps solving the concern of locating documents that are assigned low relevancy ranking.

**Question # 10: What are some of the fundamental flaws in measuring a search engine performance, and how does one overcome these issues?
Please Comment.**

- Strict standards for metadata creation might limit the number of useless or barely useful results that appear in some databases
- To overcome these issues you just have to be willing and able to take the time to learn the database/search engine you are using. In the long run it will save you a lot of time and frustration
- There is a need for more interfaces. The creator and user need to work together.
- I've seen search times which seemed respectable become unacceptably once the search is expanded to include additional qualifiers, so measuring the search speed should be done under less than ideal conditions.
- Relevance is very subjective. If several people enter the search query "IRA," their opinions on the relevance of the top results may vary widely depending on their actual information need.
- TREC has tried to address this issue. Scalability is an issue. Need to do measures with large data sets. Also, sociability and usability must be measured in any search engines' evaluation.
- Link between users. TREC test data, computer science, need human intervention. Need real questions with real users.
- Who is doing the measuring? How is the data being measured?

- I would not use recall, instead, judge performance by precision.

Summary Responses:

Table # 18: Responses from the six University Professors.

The table provides responses to the statements relating to search systems performance and measurement

Question # 09: Scholars often comment that if searchers had access to more accurate search systems, they would be more successful in their search results. Could it be that search systems are already “good enough,” so that a more accurate system would provide at best only marginal improvements?

Please Comment.

- I believe there is still scope of improving search engines
- Of course more accurate search systems would lead to more accurate searches. Well, scholars are not above asserting tautologies.
- I wouldn't expect major gains to be made, in say, expert medical searching or legal searching where vocabularies are very rigid and the users conversant in the content matter. Searching for music, however, has taken leaps forward recently, mostly by bringing old fashioned metadata techniques to the field. There are HUGE strides needed in both the multimedia and spatial worlds
- The technology is sophisticated enough now to provide good search results, but scholars still feel the systems are not good enough. The real reason is the absence of semantic infrastructure – mapping between controlled vocabulary and keywords that will point users from one to the other no matter where they start a search
- Not sure what “accurate” means----today's Google is much better than the Google of three years ago---some of this is corpus-based (better crawlers, more link structure, more documents, etc.), some is engineering based (better caches, networking), and some is search algorithm (human tuning of the SE takes place on a daily basis)

Question # 10: What are some of the fundamental flaws in measuring a search engine performance, and how does one overcome these issues?

Please Comment.

- The traditional measures of precision and recall are based fundamentally upon a notion that documents' relevance is Boolean rather than ranging over a wide variety of possible relevance strengths. A better measure would require a proper statistical model of the uses made of retrieved documents
- Not clearly defining the metrics being used and the outcome being measured. Know thy users
- The primary flaw: Assuming that one measure fits all IR contexts or tasks

Summary Responses:

Table # 19: Responses from six Information Science Organizations Participants.

The table provides responses to the statements relating to search systems performance and measurement

Question # 09: Scholars often comment that if searchers had access to more accurate search systems, they would be more successful in their search results. Could it be that search systems are already “good enough,” so that a more accurate system would provide at best only marginal improvements?

Please Comment.

- The search software itself is pretty good. The presentation of the results and the options to access the corpus need a lot of work
- This totally depends on the domain and context
- It depends on what the searcher wants. The question of what is “good enough” depends on the reason for the search
- In an ideal world, users would search for content in environments that supported various learning styles, various community practices and a full range of formats. We may never reach that ideal environment
- We can always improve a system but we may not see just how to do that today
- I believe all electronic search systems are incomplete. So, live long the books in the stacks

Question # 10: What are some of the fundamental flaws in measuring a search engine performance, and how does one overcome these issues?

Please Comment.

- Relevance, precision, and recall are each measured subjectively by a human. We assume there is only one valid answer set. I think another way to measure the results is HITS (those a human thinks are appropriate) MISSES (those a human would chose and the system did not) and NOISE (those the system chose and the system did not) NOISE can be both relevant and irrelevant depending on the level of expertise of the human reviewing the material
- Precision and recall so far as I know still require expert opinion, so there are some flaws in that process
- There have always been flaws in the process. Search engine performance (if you are talking from the results side only) are geared toward the traditional recall and precision
- I think one way to overcome these issues is to provide good help and suggestions so that people can try different “methods” of searching for the same item. It is often helpful too to ensure that both search and browse approaches are available.

A third approach is to provide different paths into the same document base (this is often done through metadata or faceted controlled vocabularies that are reflected in the taxonomy)

- The biggest problem is ambiguity of language which can be countered to some extent by controlled vocabulary and other mechanisms for refinement of queries
- However, another significant problem that is not currently being addressed is making known the scope of the content available for searching
- That search engines are accurate and all the information can be found on the web

Table # 20: Responses from two Other Libraries Participants.

The table provides responses to the statements relating to search systems performance and measurement

Question # 09: Scholars often comment that if searchers had access to more accurate search systems, they would be more successful in their search results. Could it be that search systems are already “good enough,” so that a more accurate system would provide at best only marginal improvements?

Please Comment.

- Algorithms that deal with common misspellings are useful.
- Improvements are quite possible. However, sophisticated systems that will automatically assign lots and lots of metadata tags to incoming content (this would improve results) are expensive and take a lot of expertise to set up and maintain.
- Bibliographic instruction is vital, or else people just flounder around, or think that what they find on Google or by a cursory search of ProQuest is “good enough” or even worse, that the cursory search is exhaustive

Question # 10: What are some of the fundamental flaws in measuring a search engine performance, and how does one overcome these issues?

Please Comment.

- A big problem is determining whether people find what they “really” wanted ... or even more, that they found something that they weren’t originally looking for, but that actually gave them better information than they had realized existed.
- I don’t know how one overcomes those issues—those issues existed in the days of the card catalog and the printed index.

IMPROVEMENTS IN SEARCH AND RETRIEVAL

Table # 21: Responses from 20 CENDI Member participants.

Question 13-16, relating to system improvements, data retrieval effectiveness and barriers to the user search experience

- # 13: Do you anticipate any large scale improvements in retrieval effectiveness? Explain.
- # 14: For the past 50 years or so, the challenge has been to improve the accuracy of search systems, by so doing, users will be better able to find the information that is needed. What are some of the limitations that need to be overcome for us to see more effective search systems?
- # 15: What are some of the ways to improve user search results?
- # 16: What are some of the barriers to a user search experience, and how can they be overcome or at least minimized?

IMPROVEMENTS NEEDED IN SEARCH AND RETRIEVAL:

- More efficient search systems, with improved content variety. End users will be more proficient in using the systems
- More machine intelligence built into the search tools, and the ability of the systems to learn from previous use of those systems by users
- Systems need to be more focused on the user, and more easily customizable by the users
- More parallel processing architectures and use of distributed processing
- More powerful relevancy ranking tools
- Need toolsets that can take advantage of these new resources and bring the best and most accurate information to the searcher in the shortest amount of time possible
- Better engines, better relevance ranking algorithms, and improved precision search tools in general
- Metasearch has enormous potential that has yet to be realized. Relevance ranking in a distributed environment is still in its infancy
- Interoperable categorization that can easily be understood and used by the common searcher and readily accessible training
- Better use of controlled vocabularies
- Combining improved metadata searching with natural language searching so they don't operate in isolation but are synchronous
- Figure out how to combine large data sets such as GIS and genomics data with full text and bibliographic searching, for rapid, simple-looking search
- Not enough quality metadata available for most search engines to pay attention to it, and not enough search engines look at it for data providers to invest the time it takes to create it
- Search engines ability to simultaneously search controlled vocabularies and map unauthorized terms to those authorized by the vocabularies to improve results
- Improvements in OAI harvesting and the semantic web, will have far superior retrieval than current online searching
- Testing and redesign is key to creating good user interfaces
- Improved accuracy and complexity needs to be achieved without sacrificing search speed or performance

- Natural language improvements, pattern recognition, inference, and semantic technologies improvements should occur
- The lack of comprehensive vocabularies, fully understanding diverse user requirements, simple yet powerful user interfaces, and the overall volume of non-relevant data/information are huge issues
- Process and handling large amounts of data to get decent response time.
- Need better taxonomy to improve accuracy
- Multiple thesauruses.
- Ability to drill down to get better results.
- Web search engines handle a lot of data but don't have much precision
- More metadata or the ability to search within a sentence or paragraph would help, rather than on the whole document
- Search crawlers will better understand the data they are indexing, improved searching will result
- With ever increasing CPU cycles per server, search engines will be able to derive content and content from unstructured data.
- The biggest limitation is relying on searching. You may miss the document that you are looking for because of flaws in the database
- Personalized searching might make a big improvement, i.e. the search engine is somehow intimately familiar with the types of things that you are looking for, i.e. are relevant to you.
- Powerful relevance ranking algorithms, better search interfaces, better displays of 'hits' such as categorization tools, visualization tools
- The use of categorization tools is a boost to search results
- Cataloging and indexing effectiveness play a critical role in the success of any search system's success
- It would be nice to have the capability to search all formats equally
- We need improvement in OCR (Optical Character Recognition) results
- Search engines providers will need to improve their systems in order to maintain user interest and to stay in business.
- Good tools are important.
- Tools can be improved but none will lead to a quantum leap in retrieval effectiveness
- The continual changes in languages, the differences in how individuals describes things, both in what they do and what they seek, makes it difficult for computers to full understand our thought processes
- Lots of research on different alternatives
- Process and handling large amounts of data to get decent response time. Need better taxonomy to improve accuracy. Multiple thesauruses. Ability to drill down to get better results.
- Provide alternative searching capabilities for the user to have available.
- More metadata or the ability to search within a sentence or paragraph would help, rather than on the whole document.

- There will be large-scale improvements. In the near term as more documents are created using a common XML meta tag structure and there is an metadata of imbedded into documents as they are created search crawlers will better understand the data they are indexing, improved searching will result.
- With ever increasing CPU cycles per server, search engines will be able to derive content and content from unstructured data.
- The biggest limitation is relying on searching. You may miss the document that you are looking for because of flaws in the database. Documents may not be put in the database correctly which leads to poor search results. There is a need for other mechanisms for cataloging to ensure that you have retrieved all your documents. This may require document by document review.
- Organizations that still search the bibliographic record can easily make large scale improvements to get up to the level of a Google. But for the Goggles of the world, probably “The low hanging fruit has been picked off”. Personalized searching might make a big improvement, i.e. the search engine is somehow intimately familiar with the types of things that you are looking for, i.e. are relevant to you.
- Cataloging and indexing effectiveness play a critical role in the success of any search system’s success. It would be nice to have the capability to search all formats equally.
- There is a need improvement in OCR (Optical Character Recognition) results. Because of the time element, OCR software is used to translate images to searchable text, but some words are incorrectly changed in the OCR process. More emphasis needs to be placed on quality control.
- It is great that access has improved! Search engines providers will need to improve their systems in order to maintain user interest and to stay in business!
- How information is presented to the user is important. Also, good user interface.
- By providing search instructions! Having good explanation that people can understand.
- Our Boolean search systems are accurate. The systems themselves are fine, but the data quality, interfaces, and controlled vocabulary could be improved. The algorithmic search engines used on the web also seem to be “accurate” enough for their purpose, which is only to find approximate results
- Improve the interface design and data quality. Full text databases have been improved by adding fields and controlled vocabulary. But I don’t believe the reverse is often true. Adding full-text does not increase relevancy of results.
- More customization
- Better understanding of users’ needs and abilities and ways of searching. Cognitive psychology.

BARRIERS TO OVERCOME TO IMPROVE USERS SEARCH EXPERIENCES:

- The biggest “barrier” for me is the amount of information available
- A lack of understanding by users regarding how information is published electronically and rampant inconsistency in the construction of data and indexes
- I would think that effectiveness is “in the eye of the searcher” and if we get good marks from our customers for our systems, that’s the most important gauge
- Language; bad presentation; poor technology
- Ways to overcome – usability testing, usability testing, usability testing
- Lack of user testing often leads to problems with the finished product. Often what seems logical and obvious to designers is not clear to users
- Users are impatient and unwilling to scroll to information that is not visible on the first screen
- User education, both formal and informal, is a key way to improve user satisfaction and users’ search capabilities
- I think the “suggested terms” for users probably does improve user satisfaction with their search results
- The quality of the data affects the quality of the search experience
- Withheld or buried information – the absence of explanations of how the search system works is the greatest barrier to a user’s search experience
- Better search tips/help files
- Better metadata describing the content of the records
- One barrier is poor design of the website or database
- Content sensitive. Make things simple.
- Speed would be a benefit to increase response time. Processor speed is holding us back.
- Home bandwidth has limitations to the user. This places limitation on downloading capability
- Need for time, patience and knowledge
- There is a need for other mechanisms for cataloging to ensure that you have retrieved all your documents. May need to do document by document review.
- Improvement in user interface will also improve the user experience
- Metadata search systems with interactive controlled vocabulary to improve search results
- We need more training for users of the various systems.
- Users could benefit if search systems had the help information up front and readily available to aid users with each section
- Suggestions on the hit list or the ability to refine searches would also be good
- Providing search instructions and having good explanation that people can understand.
- Giving the user the ability to place their idea into a search experience.
- Users obtaining support from intermediaries, such as the library, for assistance.

- Improvements in data quality, interfaces, and controlled vocabulary can greatly enhance a searchers experience
- Multiple thesauruses and drill down capability.
- Content sensitive. Make things simple. Speed would be a benefit to increase response time. Processor speed is holding us back. Home bandwidth has limitations to the user. This places limitation on downloading capability.
- Need for time, patience and knowledge: Intermediary. Time and patience! Familiarity with subject and collection.
- There is a need to segment the collection to improve search results, for example, in subject categories. There is also a need to apply a broad thesaurus across specific categories of content. This will provide searchers additional clues and options.
- There are not enough human factors in building interface. Developed for good searchers, but not designed for the novice searcher.
- Language will become an issue as the percentage of Americans speaking English as their primary language declines.
- There is the need for more powerful relevance ranking algorithms, better search interfaces, better displays of “hits” such as categorization tools, visualization tools, etc.
- Users have great expectations that whenever they do a search, that the result will be more “Google like.” Google search results have become the standard by which user expectations are based. Improvement in user interface will also improve the user experience. The use of categorization tools is a boost to search results.
- Some metadata search systems do not display their controlled vocabulary where it is obvious to the user to improve their search results, by being interactive; instead it is left to the user to determine that there is such a tool. This can be counterproductive when one considers the time, cost and effort in maintaining a controlled vocabulary.
- From a searching capability, users can improve their skills. Post processing.
- We always need more training for users of the various systems.
- The user’s inexperience, lack of knowledge, and lack of training. Users could benefit if search systems had the help information up front and readily available to aid users with each section. Suggestions on the hit list or the ability to refine searches would also be good.
- Most users have time constraints. If the search experience is difficult, then users will move on. Good tools are important. Users will get frustrated if results are hard to find.
- Giving the user the ability to place their idea into a search experience. Users obtaining support from intermediaries, such as the library, for assistance.
- The two biggest issues for the user experience are the interface design and search engine transparency. If the interface is designed well, even a novice searcher can take advantage of features that in another system would be considered advanced or complex. A major flaw of internet search engines is their lack of transparency. The user doesn’t know how their search is being interpreted, executed or sorted.

- I believe individuals can do their own searching.
- Convince people that they can get information. Get what they are looking for.
- If the system become more interactive, then searchers will get better results
- A lack of clarity in the mind of the searcher.
- Take time to learn how system works, ask for help from an expert, try multiple search engines, and use controlled vocabulary...
- Language, fear, general state of mind, mental illness, physical distractions, attitude of user, design of search system, physical disabilities, level of education, etc.

Table # 22: Responses from 14 DOD Organizations and DOD Contractors participants.

Question 13-16, relating to system improvements, data retrieval effectiveness and barriers to the user search experience

13: Do you anticipate any large scale improvements in retrieval effectiveness? Explain.

14: For the past 50 years or so, the challenge has been to improve the accuracy of search systems, by so doing, users will be better able to find the information that is needed. What are some of the limitations that need to be overcome for us to see more effective search systems?

15: What are some of the ways to improve user search results?

16: What are some of the barriers to a user search experience, and how can they be overcome or at least minimized?

IMPROVEMENTS NEEDED IN SEARCH AND RETRIEVAL:

- The modern element of information retrieval impatience is one problem that needs to be addressed
- The deep net/hidden net needs to be more fully explored and better ways developed to utilize information hidden there
- If some standardization were possible to be achieved in the industry, everyone could be reading from the same page of music.
- I think serious scholarly researchers would like the twin internet system
- Insist on metadata for all documents on the web
- A means to create a search system with taxonomy. Users not thinking in terms of the way subject headings were created.
- More natural language
- More acceptances in satisfying the common user
- Ability to combine and manipulate search sets
- Ability to review the search results in a bit more detail - on screen output could be designed differently (dynamically) from the output formats used to generate bib files
- Allow searching using limited distribution; adjust searching in an advanced mode to allow for extended searching in the various volumes

- Facet search results...coming from results sets...takes author name associated and group it. Also clustering.
- Expect improvement in audio and video, they are both poor today. The need is there. Also, language processing, real shift 10-15 years, as performance systems improve. TREC language processing to improve retrieval has not resulted in improvement.
- There will always be the issue of human interaction that breeds inconsistency, for example in indexing. With machine aid indexing will reduce cost and time, but it will not be as good as human beings.
- Usability issues. Systems do not interact well, documents versus multi-media! Inaccuracies in search systems. Cataloguing is a problem.
- We have not yet figured out the most effective search interface. There is the need to help the user formulate searches for better results. A need for commonality across search systems to allow for the exposure on information space to improve search results. Need usability of system with post retrieval exposure.
- Ability to do a broad search, select a subset of the broad search and then print out the selected records and the non-selected records in two different bibs

BARRIERS TO OVERCOME TO IMPROVE USERS SEARCH EXPERIENCES:

- Teaching better strategies
- “For Dummies” help tips that are tested by young searchers
- Training. Education.
- Users do not know how to select useful and relevant search terms.
- Users do not understand how to search
- Users do not learn how to use various databases
- Lack of knowledge to controlled vocabulary is a hindrance
- There should be an online thesaurus for any database with controlled vocabulary
- By making searching easier
- If accurate metadata is assigned, then search results should improve.
- Better training – librarian-developed and provided.
- Too much available – users are confused
- Give them more fields to add terms and narrow the search to make it more focused
- It’s frustrating when searches time out.
- It’s frustrating when a complicated search strategy fails but then cannot be recovered for review and tweaking
- Helpline with a human versus a computer/recording
- IT developed interfaces are the number one barrier. They are not intuitive to the average, or even sophisticated searcher.
- A better understanding of the role that information seeking behaviors play in the success of search experiences is critical. A holistic view that does not just focus on the search engine’s results but instead looks at the whole user experience will make the difference.

- Understand the user population (How many are novices or new to the website, which will determine how familiar they might be with the navigation and the terminology used on the website? What are they seeking when they come to the website? Do the majority need high recall or high precision?).
- Search terms may not match jargon or business-specific terminology. Content may include a number of synonymous terms, depending on the author, where uniform use of terms would be better.
- Acronyms used in the content may not be familiar to users.
- User's unwillingness to specify searching needs. Poor user selection of search terms! Lack of experience in searching. Through training and education a searcher experience will improve.

Table # 23: Responses from six University Professors.

Question 13-16, relating to system improvements; data retrieval effectiveness; and barriers to the user search experience

13: Do you anticipate any large scale improvements in retrieval effectiveness? Explain.

14: For the past 50 years or so, the challenge has been to improve the accuracy of search systems, by so doing, users will be better able to find the information that is needed. What are some of the limitations that need to be overcome for us to see more effective search systems?

15: What are some of the ways to improve user search results?

16: What are some of the barriers to a user search experience, and how can they be overcome or at least minimized?

IMPROVEMENTS NEEDED IN SEARCH AND RETRIEVAL:

- Yes, with the use of large computing power available and innovation in parallel algorithms
- Lower precision
- Incorporate semantic searching.
- Improve precision and classify result sets
- For internet-wide searching, probably not, as there appears to be no prospect for moving people away from WYSIWYG visual formatting of documents to logical/structural markup
- Incorporate more secondary sources or information
- Many of today's search algorithms were developed in an environment of computing scarcity. Now we can play with more inductive and heuristic systems
- The ability to cross domains in searching and better synthesize results...to get some picture of how 'ALL' the documents fit together
- The combination of internet search engines and commercial databases and library OPACs will probably improve information retrieval on a larger scale than ever before

- Improved retrieval effectiveness...passages; multimedia; cross-language

BARRIERS TO OVERCOME TO IMPROVE USERS SEARCH EXPERIENCES:

- Listen to the user
- Interfaces can get better, and more integrated into user workflows
- Developing ontology's for domains and mapping keywords and controlled vocabularies
- There are different user groups and their information searching literacy levels vary greatly
- Educate the users. Information literacy should be incorporated to school/university curriculum
- When people become information literate, the barriers will be minimized
- Get people to use relevance feedback

Table # 24: Responses from six Information Science Organization Participants.

Question 13-16, relating to system improvements; data retrieval effectiveness; and barriers to the user search experience

13: Do you anticipate any large scale improvements in retrieval effectiveness? Explain.

14: For the past 50 years or so, the challenge has been to improve the accuracy of search systems, by so doing, users will be better able to find the information that is needed. What are some of the limitations that need to be overcome for us to see more effective search systems?

15: What are some of the ways to improve user search results?

16: What are some of the barriers to a user search experience, and how can they be overcome or at least minimized?

IMPROVEMENTS NEEDED IN SEARCH AND RETRIEVAL:

- Better application. Adding controlled terms and allowing use of all synonyms in search
- Providing several ways to search so that most learning styles and cognitive processes are accommodated
- Use the controlled vocabulary to expand search queries as well as to apply metadata to the records; use it at both ends
- Problems have to do with different uses of the same terms by different groups
- Vocabulary control and subject switching – not new. Better semantic understanding.
- Ontology's when implemented through semantic web tools will help to make certain types of information more retrievable

- Retrieval effectiveness will be helped by further development of portals, customized environments and retrieval systems that “learn” from the users experience what he or she wants
- The ability to search for not only terms but how they relate to one another
- We need good tools to turn our current tools, like thesauri, into richer structures
- We need subject matter experts to help in these areas as well, since they can also help by building these structure in the front-end
- Incremental improvements can be made based on how users behave in their information seeking tasks
- Systems will be improved slowly as the creators of those systems get a solid sense of what people are trying to do in the online environment
- Our effectiveness is limited by our factory-style approach towards search [one-size-fits-all]. We don't build systems that accommodate a wide variety of contexts, learning styles or formats.

BARRIERS TO OVERCOME TO IMPROVE USERS SEARCH EXPERIENCES:

- Accommodating the vernacular is the big problem that is disambiguation of terms effectively and of course allowing different ways of access to the data
- Presentation of results, manner in which search is allowed. These are not really hard changes to make. We have the tools at hand
- Clearly user education although that is very hard
- Clustering and visualization is the future to help people hone searches
- One of the biggest barriers to a user search experience is the lack of time a user is willing to spend on a search
- Users who don't know what they don't know.
- User education is a must. The user has to understand how the system functions (at least to some extent)
- Systems will have to be capable of recognizing instances where help might be useful
- Inconsistent human intervention. You have to have systems that allow for inconsistencies of human behavior
- Limited knowledge of what to expect from the system

Table # 25: Responses from two Other Libraries Participants.

Question 13-16, relating to system improvements; data retrieval effectiveness; and barriers to the user search experience

13: Do you anticipate any large scale improvements in retrieval effectiveness? Explain.

14: For the past 50 years or so, the challenge has been to improve the accuracy of search systems, by so doing, users will be better able to find the information that is needed. What are some of the limitations that need to be overcome for us to see more effective search systems?

15: What are some of the ways to improve user search results?

16: What are some of the barriers to a user search experience, and how can they be overcome or at least minimized?

IMPROVEMENTS NEEDED IN SEARCH AND RETRIEVAL

- Better relevance ranking algorithms need to be developed to enable better precision in full-text searches.
- Lack of precision in results algorithms; lack of (affordable) software to automatically categorize incoming content in databases; lack of customized, individual taxonomies and controlled vocabularies. LCSH is not a “one size fits all” controlled vocabulary.
- Improved controlled vocabulary, misspelling corrections, lots of searchable field limits.

BARRIERS TO OVERCOME TO IMPROVE USERS SEARCH EXPERIENCES:

- Get some training from an information professional on how to construct better searches; doing some digging on the database to see how the content is organized and what thesauri or metadata is used on the database, and then use those terms in conjunction with full-text searching.
- The refusal of users to consult librarians and other informational professionals is maddening. That barrier can be self-generated, or perhaps the user has had unpleasant experiences with librarians.
- Lousy database design—unhelpful help screens and “term not found” notices, with no mechanism for bumping people back to the original search page, or no mechanism for suggesting other terms to use if the ones they use aren’t in the database.
- We are subject to too many market pressures—making our dbs like Amazon or Google.
- Education, clearly and simply written help screens.

FUTURE ROLE OF CATALOGERS AND INDEXERS

Table # 26: Responses from 20 CENDI Member Agencies Participants.

Question 25 & 26, relating to the future role of catalogers and indexers

25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?

The respondents were mixed in their views as to whether there is still a role for catalogers and indexers in support of the quality of search results. The following reasons were provided:

- Catalogers and indexers role remains and will increase in the future
- Metadata is essential to the management of the full-text data over time
- With full text searching, there are more opportunities for humans to apply metadata to documents and improve search results
- Catalogers and indexers should still play a central role in highlighting the key concepts, topics, names, places, etc. that are found in a document or record
- Their role is still a needed support
- If automation can help their process, that improves the overall process significantly
- Improvements in search results can be obtained, over full-text, if high quality, skilled, and domain expert catalogers exist.
- Indexers and catalogers play an important role in providing quality metadata
- These disciplines were once essential, but their roles have changed to being helpful but not essential
- Both groups will ultimately be eliminated as non-essential expenditures
- If catalogers and indexers are providing a quality product, then they are enhancing searching
- Catalogers are still needed for descriptive metadata.
- Human catalogers will play a lesser role with respect to subject cataloging
- I don't believe the role of catalogers is minimized. You still need descriptive metadata.
- Where indexers are still needed is in coming up with terms not in the actual text, synonyms or a concept talked around but not mentioned.
- Don't believe the role is minimized. The role needs to be automated.
- Catalogers will play a lesser with respect to subject cataloging, but still important role.
- Catalogers and indexers must maintain a high level of quality to output.
- Indexers are still needed to identify terms excluded from the text, synonyms or a concept talked around but not mentioned
- Indexers are also needed to add terms later for new names for concepts and changes in author names
- Catalogers and indexers are even more important as the size of our collection increases
- Catalogers are needed to ensure accurate input
- Good catalogers are needed to ensure accurate input, so that documents are accessible to searchers.
- Catalogers are needed to improve descriptors and identifiers for effective research and retrieval
- Their expertise is necessary to get to the relevant documents
- Failures of full-text searching show the need of catalogers and indexers
- The failures of full-text searching show the need of catalogers and indexers. That is why the web now uses metatags, to try to get people to catalog their own works.

- In an ideal world, catalogers and indexers working together with developers could make a better system.
- Most catalogers see their work as an art form. They are not connecting people to information.

26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?

The respondents overwhelmingly support a role for human intervention in metadata indexing to improve the quality of search results. The following reasons were provided:

- Perhaps someday, human intervention will not be necessary, but so far, irrelevant terms still need to be removed and relevant terms still need to be added
- The role may be limited, and more as a quality control function
- A need exist when dealing with classes of information such as numeric data, images, software, charts, audio and multimedia files
- Human created metadata will enable better searching, but it is becoming unaffordable
- Even the best automated metadata indexing requires management and the introspective review
- The need exists, especially if multiple languages and data types are to be searched
- Metadata indexing (done by people) is important to improving the quality of search results for experienced and “power” users.
- Machine aided indexing at least provides consistency
- There is always a need for human review to ensure accuracy
- Human intervention is needed to ensure good quality control.
- Human intervention is critical
- It is always important to have human intervention to improve the quality of your results.
- Machines will ultimately be able to suggest all metadata, but there will a few pieces of metadata that you will always want human review to ensure that it is accurate.

Table # 27: Responses from 14 DOD Organizations and DOD Contractors participants.

Question 25 & 26, relating to the future role of catalogers and indexers

25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?

- It depends on the application and how people will be looking for information.
- Should not be. People do not understand the role of librarians. There is the perception that their role is minimized with the advent of full text searching.

- Users still do not know the importance of using synonyms, Boolean techniques or how to tweak their results to increase or decrease the quantity, relevance or accuracy

26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?

- I don't believe that metadata indexing improve search results except for bibliographic metadata such as date, title, author.
- There is still a need for human intervention.
- Yes. Humans first, then machine next. The final decision should be made by human.
- There will always be a need for catalogers/indexers to get the correct metadata into the file;
- Electronic metadata creation cannot achieve complex analysis
- Only the human mind can make the necessary distinctions with natural word syntax, language idioms and slang's which greatly affecting searching capabilities

Table # 28: Responses from six University Professors.

Question 25 & 26, relating to the future role of catalogers and indexers

25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?

26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?

Respondents support the view that there is still a role for both catalogers and indexers, maybe not minimized but more of a shifting in responsibilities. The role of catalogers in supporting searching is believed to have changed, however, other roles such as collection management is thought of as important and valued. Other views expressed included the following statements:

- While metadata may be generated automatically, there is still a need for human catalogers to verify content
- The maintaining of controlled vocabularies, the creation of new ones, and mapping between keywords and controlled vocabularies will occupy more of catalogers time

On the other hand, the view is supported that catalogers of the future will not be assigning terms to objects, but instead, they will be tuning data mining algorithms. Human intervention is needed for acquiring or creating the metadata.

Table # 29: Responses from six Information Science Organizations Participants.

Question 25 & 26, relating to the future role of catalogers and indexers

25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?

The majority of participants agree that catalogers and indexers roles have changed. The explanation provided for these changes varied however. The following summarized these views:

- While the role has changed, full text alone is not the answer
- The question is more of balance and how much manual intervention is appropriate
- Increasingly the machine becomes more the doer and the human becomes the quality controller
- The subject matter expert is important to ensure the search algorithms stay honest
- Catalogers and indexers may do less actual metadata creation, but they do more knowledge base and rules development
- Intelligent indexing by those with a knowledge of the field is a highly desirable if costly value

26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?

There was full support expressed for the need for human intervention in metadata indexing. The views are summarized below:

- While terms can be gathered automatically, a review is needed for those that are incorrectly presented
- Manual oversight and intervention is necessary
- long as language is ambivalent in its usage there will be a need for human intervention
- Another set of eyes and another brain in assessing and evaluating the content is always desirable

Table # 30: Responses from two Other Libraries Participants.

Question 25 & 26, relating to the future role of catalogers and indexers

25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?

- If catalogers are not constructing taxonomies or thesauri, or assigning metadata terms to be used with the databases, it is very easy for administrators to say that catalogers and indexers are no longer necessary.

26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?

- Absolutely.
- Of Course.

IMPROVING SEARCH RESULTS...ROLE OF METADATA AND FULL-TEXT

Table # 31: Responses from 20 CENDI Member Participants.

Question 19, 20 & 21, relating to improving search results.

19. What role should metadata play in improving search results?

- Results could improve if the metadata itself improves and is more widely available for all data.
- Only if it describes numeric data, images, software, charts, audio and multimedia files
- Well constructed metadata will improve search results
- It affords a structure for the development of the consistency necessary for a user to confidently construct quality searches
- A very critical role, however it is highly underutilized
- Must be there in the content and may also be used for advanced searching
- For those who like browsing, controlled vocabulary is a must
- Metadata makes search results more relevant
- Metadata can greatly improve the information we want to identify
- Helps in narrowing the search.
- Meta tagging should complement full text searching
- Metadata can play a role in the categorization
- Data needs to be input correctly and accurately. There needs to be consistency!
- Controlled Vocabulary is important.
- The same role it currently plays in most databases
- Helpful in getting searchers to the information they need, but must understand how the system works.
- The more metadata the better.
- Role? Perhaps the question needs rephrasing: To what degree should metadata improve results? Again, it depends on the quality of the metadata. "Garbage in, garbage out."

20. Does full-text searching eliminate the requirements to construct metadata?

- Full text searching makes it possible to search without metadata.
- Metadata helps enable searching
- Absolutely not, it is an opportunity to enrich the text

- Visible metadata and controlled vocabulary can help researchers retrieve all of the results that include that precise term
- It depends on what you are trying to do. Labor cost. It is intensive
- No. I don't believe so. You need both the descriptive metadata, such as author, title etc., and the subject metadata for synonyms to the words in the article.
- No. Full text searching demands adding meta tagging to make the searching useful
- No. Still need metadata for classification / limitations on the document, etc
- No. There still has to be metadata.
- No. That is why we now have meta tags in html
- No. There still has to be metadata. There is a difference between digitization and preservation. There is the need to preserve the metadata or descriptions near the files described.
- Probably not.
- Does it or should it? Yes it does, no it should not.

21. Can full-text search be used to effectively augment metadata?

- Yes, since a mix may be the optimal approach
- Metadata has very limited use except for numeric data, images, software, charts, audio and multimedia files
- Yes, the two are often used together successfully to produce more precise search results for users at various levels
- Yes.
- Yes. Will get some false hits too.
- They should complement each other
- Yes. Though I believe the more relevant statement is that metadata can be used to augment full text searching
- Yes. The two work together well.
- Sometimes, if the metadata quality is poor
- Yes. And vice versa

Table # 32: Responses from 14 DOD Organizations and DOD Contractors Participants.

Question 19, 20 & 21, relating to improving search results.

19. What role should metadata play in improving search results?

- It is important as a primary level
- The basic and most critical elements are defined
- Metadata should assist in improving search results
- Every document in a database should have metadata/bibliographic data for search retrieval
- Improve the accuracy

- Good metadata is not widely supplied with government information and searching by metadata requires you to know the appropriate government jargon to match.
- It is good in refining search results. The biggest problem is getting people to know and learn so as to improve search results.
- Metadata should automatically map to content those terms are important to search results.
- Useful in improving relevancy (Recall). Limiters, for example, the first 500 hits.
- An increasing role, depending on the database. Metadata is not necessary with photographic databases.

20. Does full-text searching eliminate the requirements to construct metadata?

- Absolutely not.
- If results are less than expected, metadata might be the only other way to extract the data
- No. Metadata is essential for bibliographic information—author, title, date, etc.
- In Google search, metadata searching is not needed. With the searching of pharmaceutical databases for example, a 80/20 precession/ recall does not cut it. For specific collections, metadata is needed.
- No. Not if the working world must go on. There are only so many hours in the day.

21. Can full-text search be used to effectively augment metadata?

- Perhaps
- Yes, first find results based on controlled vocabulary/bibliographic data/metadata, then search by full text for specific items
- Using vague search terms can retrieve too much information or low relevancy
- Unless you are searching for a specific author or title or date, full- text search is essential.
- Yes. Product names.
- Yes. After one has exhausted metadata searching, full text searching is a second choice.
- Yes. They augment each other. There are drawbacks in standard full text searching. Words are full text. This does not take care of homogeny.

Table # 33: Responses from six University Professors.

Question 19, 20 & 21, relating to improving search results.

19. What role should metadata play in improving search results?

- Significant

- Your question suggests an idea that metadata is an end in and of itself. I see it instead as a means
- Offering fielded search, organizing search results by categories, and displaying results in a consistent look
- Metadata is one of many sources of evidence for searchers, it is too expensive to produce manually

20. Does full-text searching eliminate the requirements to construct metadata?

- Metadata may often have other non-searching roles to play
- No, generate as much as you can automatically

21. Can full-text search be used to effectively augment metadata?

- To augment the metadata itself? Certainly.
- Of course---that is what people want anyway

Table # 34: Responses from six Information Science Organizations Participants.

Question 19, 20 & 21, relating to improving search results.

19. What role should metadata play in improving search results?

- A big one. Control of the terms in use and the way they are applied is crucial
- Structured controlled vocabulary to get to concepts
- It should be consistently reliable and made searchable according to the need of the user
- Cross-referencing

20. Does full-text searching eliminate the requirements to construct metadata?

- No, it makes it more important due to all the false drops from using the same term in different meanings and as pictures speech
- It depends on users, context and objective of system
- No. Having both is best
- Metadata is the way to provide links between documents
- Metadata is the way to add information that does not appear or is not readily discernable from the document itself
- No, it is imperative that we have **both** in place so that users with different through processes and learning approaches can be successful in their information seeking

21. Can full-text search be used to effectively augment metadata?

- Yes – I think that was the original idea
- Absolutely
- It would be lovely to be able to turn it off and on as need dictated

Table # 35: Responses from two Other Libraries Participants.

Question 19, 20 & 21, relating to improving search results.

19. What role should metadata play in improving search results?

- It's vital.

20. Does full-text searching eliminate the requirements to construct metadata?

- Absolutely not. Searching “automobiles” won’t find documents that are about “cars” unless a taxonomy/thesaurus schema is running in the background to guide people

21. Can full-text search be used to effectively augment metadata?

- Oh, of course. I use full-text searching all the time (on Google, that’s all there is).

CONCLUSION

The purpose of the study was to review the status of search methodologies and the issues and concerns that affect information seekers search results. An attempt was made to answer the following questions. What are some of the current and desired searching capabilities? What are some of the limitations and barriers that need to be addressed in order for information seekers to find answers to their questions or to seek new knowledge? What are the preferred methods of searching and the rationale for these decisions? What role will catalogers and indexers play in the future? Finally, what will search in the future provide that is currently not available?

The 48 participants represented 29 organizations and federal agencies from a cross section of information science professionals at various levels of responsibility. They included: senior level scientific and technical information managers; university professors; university librarians; federal agency librarians; information science providers; information science instructors/trainers; and other information science professionals.

The conclusions drawn from this study are based on the following assumptions. All participant responses are given the same weight regardless of their organizational status. There was no differentiation made based on level of technical skills or knowledge.

A summary of the findings from participants responses were grouped into seven categories.

- *Preferred method of searching*
- *Status of searching methodology*
- *Limitations in using full-text and metadata searching*
- *Measuring search systems' performances*
- *Improvements in search and retrieval*
- *Future role of catalogers and indexers*
- *Role of metadata in improving search results*

These responses were incorporated where possible with the views from the reviewed literature.

The first category sought participants' view on the **preferred method of searching**. While there were a few participants who distinctly preferred full-text searching and a few who preferred metadata searching, the majority of participants used both methods to search. The third group's preference often varied based on their knowledge of the subject being searched, the richness of the database, and the type of question being asked.

Participants who favored full-text searching gave the following reasons for their choice; the ease of use and speed achieved with full-text search engines. Also, the incompleteness and inaccuracy of metadata was a disincentive for metadata searching.

Participants who favored metadata searching gave the following reasons for their choices. They felt that this method of searching was more suited for searching government information. It also provided better results with good response time and high precession.

Participants who favored both methods of searching gave the following reasons: Their method of searching was dependent upon their knowledge of the subject being searched, the richness of the database, the comprehensiveness of the database, and the type of information that was being sought. Some participants conducted an initial key-word search with a follow up metadata search to improve or refine their results.

The second category related to the **status of searching methodology**. This was an attempt to solicit information as to where we are in searching methodology and where the field is heading.

Some participants believed that it is important to provide an access system to accommodate both full-text and metadata searching. This would be achieved by having a rich mixture of metadata and taxonomies with crosswalks between them to enhance search results. The goal is to keep the widest range of approaches available so as to give information seekers flexibility in conducting searches. Another view is to place greater emphasis on identifying the user community so as to build interfaces that will accommodate their specific needs. The user community would specify what it is looking for and then algorithms would be built to fill search limiters that were previously found useful. Automatic metadata generation and indexing is also advocated. The belief is held that taxonomies and metadata are ways to reduce noise in search. It was also suggested that there is an on-going challenge for content providers to develop specialized collections that meet the needs of specific communities of practice.

Other thoughts included the application of improved structure to the data to enhance search results. The mixing of full-text and meta tagging is supported in order to improve search results, providing the taxonomy is consistent and is also consistently applied.

One may conclude that the shared thoughts surrounding the status of searching methodology support the view of a blurring of the two prominent methods of searching.

The third category addressed some of the **limitations in using full-text and metadata searching**. Each search methodology is discussed below.

Participants views regarding **limitations to full-text searching**:

- Less relevancy in search results
- Overwhelming volume of search results
- Too many hits and term ambiguity
- Results do not match searcher's query
- Limitation in information layout
- Lack of synonyms...inability to differentiate
- Lack of precision and control

Participants views regarding the **limitations in metadata searching**:

- Information seeker at the mercy of the metadata creator
- Expensive to create and maintain
- Too much information noise...may miss important information
- Ambiguity of terms, too few hits and missing categories
- Inconsistency in data presented
- Unfamiliarity with metadata rules may result in poor output
- Requires more education and thought

The **fourth category** relates to **search systems' performance** and the ability to **effectively measure performance**.

The overall responses support the need for improvement in search engines ease of use to support information seekers' customization of search results. Improved search tips and help guides to support more effective searches are also advocated. Participants believe that there is a need for improvements in interface design and usability to promote more seamless search systems. Also, there is the belief that with improved algorithms, more effective search results will be realized.

On the input side of the system performance, it is the view that by improving catalogers' application of the controlled vocabulary through training, information seekers will obtain more effective search results.

Participant views on the **fundamental flaws in measuring search systems' performance** include:

- Lack of virtual support and education for less experienced searchers
- Search engine performance tends to be measured by user expected search results
- Lack of strict standards for metadata creation gives poor search results
- Lack of clarity in the metrics being used and the outcome being measured
- Tendency to judge search engines based on speed rather than accuracy of results
- Measuring based on the number of hits
- Using information seeker satisfaction as a measure of success
- Scalability is an issue...need to measure large data sets

The literature review on the current and desired searching capabilities supports some of the views expressed by the participants in their comments noted above.

Mike Moran (2006) suggested that a "good search engine" should not be one's goal; instead, searching can be viewed as a means to an end.

The ease of use of the vast number of broad base search engines such as Google, Yahoo, MSN, come with a price that too often information seekers find unacceptable, that is, the vast number of hits that perhaps fail to include the correct answer to ones question. Outsell in a 2006 survey,

estimated that broad base search engines only account for approximately 68% success rate. These high search failures bring with them an economic cost. In a January 2007 survey of broad based search engines by Hitwise, it was reported that Google accounted for 63.1% of the ten million searches over a four week period. Yahoo accounted for 21.4% and MSN 10%.

Oliver Scheffer (2007) reported that broad base search engines were missing the most valuable part of the web, often referred to as the deep web, or the invisible web. Listed below is Scheffer's estimate for deep web search sites by content type: ***Special Databases 54%; Internal Databases 13%; Publication Sites 11%.*** These three content types account for approximately 74% of all deep web sites. The result shows that deep web content is accessible to selected information seekers and is not available to the general public.

The author advocates "fully customized vertical search." These search engines can address the informational needs of specialized or focused audiences and professions. Such search engines may be designed to support job seekers, doctors, engineers, scientists, or lawyers. They are able to deliver more relevant and essential information than is found with broad base search engines.

It is therefore important to differentiate the limited capabilities of broad based search engines to those of vertical search engines when the goal is to seek more knowledge on a specialized field of study. The notion that "Google has everything" is a fallacy. For general searching, broad based search engines such as Yahoo, Google and MSN can be a source for quick answers to the information needs of the general public, but their sources of information fall short in meeting the needs for specialized disciplines where access to the 'deep web' is critical to a researcher or scientist.

Tom Reamy (2006) supports faceted navigation as an alternative to search and browse. The author noted its dynamic capability of combining searching and browsing of compound subjects in an intuitive process.

Andrew Pace (2007), suggested that the information seekers' experience can be improved or enhanced by "making the bibliographic data work harder for the user through the establishment of relationships between the bibliographic data and other systems." The North Carolina State University faceted browser interface is an example of such a system.

Enhanced gateways are also recommended as a way to improve the information seeker experience by centralizing or simplifying the search process. Pace (2007) referred to institutions and organizations linking bibliographic data from other sources in an attempt to find a book or a document from a local library, or to find associated bibliographic data about it, that can be purchased online through a link to a retailer.

The **fifth category** examined the participant views regarding the improvements needed for search and retrieval effectiveness and the barriers that need to be overcome to improve information seeker experiences.

A summary of the recommended **improvements in search and retrieval** include:

- More focus on the user with customized capabilities
- Synchronization of natural language with improved metadata searching
- More effective testing and redesigning to provide effective user interfaces
- Improved parallel processing architectures and the use of distributed processing
- Improved retrieval effectiveness...passages, multimedia, and cross-language
- Ability to cross domains in searching and better synthesize results
- Ability to search based on terms and their relationship with one another
- Ability to search within a sentence or paragraph
- Processing and handling large amounts of data
- Ability to search all formats equally
- Usability issues...systems do not interact well...documents versus multi-media

A summary of the **barriers to the user search experiences** include:

- Too much information available
- Lack of user testing
- Inadequate and inaccurate metadata describing data content
- Information seekers poor searching skills
- Lack of interfaces that are integrated into user workflow
- Inadequate ontology's for domains
- Inadequate mapping of keywords and controlled vocabularies to ontologies domains
- Inadequate use of clustering and visualization tools to improve search results
- Information seekers unwillingness to allocate more time on a search
- Home bandwidth limitations
- Lack of processing speed
- Need more metadata search systems with interactive controlled vocabulary
- Improve the information seekers ability to place their ideas into the search experience
- Lack of data quality

A summary review on the limitations and barriers to overcome to improve the information seekers' searching capabilities reflect some of the thoughts and ideas presented in the participant's responses as stated above.

Maybury (2005) suggested that in order to make search and discovery work, it is critical that both the barriers and solutions to search improvement methodology must be addressed. The author advocates a holistic or systems approach to successfully address these issues.

What is required? Conduct a technological assessment whereby the search system capabilities and activities are analyzed. One must develop a clear understanding of the barriers to retrieval if improvements in searching capabilities are to be attained. How is this achieved? The following are recommended: develop an understanding of the information seekers' behavior; understand

their query intent versus query results; understand their navigational capabilities; a shift in focus from defining metadata to analyzing usage; optimizing search locally; engaging vendors; and infusing practice with system engineering rigor. Both Kerchner (2006) and Nicholson (2004) also advocated a holistic approach for improvements to be realized by the information seeker. Performance measures must include the total search experience. The information seeker search experience begins when an individual enters key words in the search box, through the search system feedback with results. The search success should be based on the information seeker's assessment of the usefulness of the information retrieved.

A system evaluation is critical to the success of any search system. Such an evaluation must include: usability testing and assessment; user satisfaction surveys; review and assess search logs; analyze system response time and downtime; review captured information seeker queries; incorporating frequently used search terms that are excluded from the search systems controlled vocabulary so as to enrich the search experience. Kerchner (2006) noted higher precession levels when the IRS incorporated captured queries into their controlled vocabulary.

Hawkins and Thomas (2005) estimated that there are some 17,000 government websites in which a large portion lack search interfaces. This makes searching a challenge. The authors propose a hybrid approach to data access, whereby distributed and centralized techniques are applied.

In Carol Tenopir's presentation at the 2005 "Search Engine Meeting" in Boston, she noted that there is a clear distinction between students and experts search experiences. Students search internet search engines over formal electronic sources as their first choice. The focus is on simplicity and speed. Expert searchers on the other hand, do both browsing and searching, with usage patterns varying with the subject being searched. Tenopir view therefore supports the importance of developing an understanding of one's information seeker behavior in order to provide effective search tools and resources.

While scholars have emphasized the importance of establishing "best practices" in the designing of user interfaces, Resnick and Vaughan (2006), caution that designers are faced with the dilemma of trying to appease two types of searchers, the expert and the novice. The authors note that the novice searcher has little or no interest in learning the rules (developing an understanding of each search engine architecture and algorithms), while the expert searcher is more willing to accept the challenge in his quest for inquiry. Karen Markey (2007) also reported in part one of her research findings on end-user behavior that "end-users do not resemble the systemic approach of expert intermediary searches who use the Boolean OR operator to build intermediary sets of retrievals for the unique facets of user queries." She noted that end-users apply a few short search statements, usually two to four words in their search strategy. Their gratification comes from doing their own searching where it is convenient, immediate, and instantaneous by linking to the internet with the hope of retrieving full length documents from their subject search. Bishop et al., (2000), Cooper (2001), Jansen (2005), and Markey (2007), all agreed that the only advanced features that appeal to these users are quotes for bound phrases and plus (+) and minus (-) operators. When advanced search system features are used, they are likely to be used incorrectly. Markey (2007) noted, from her 25 years of end-users research

findings, that this group of searchers does not take advantage of available search tools. Their search strategies make information retrieval appear to be a very simplistic process.

Resnick and Vaughan (2006) noted that system design should be treated as an iterative process, with continual review to seek improvements and fine tuning as information seekers needs and demands change. The emphasis here is to develop an understanding of the user community behavioral patterns and adapt to the changes necessary in order to remain an effective information provider whose data will remain appealing to customers, and sought by new ones.

Karen Markey (2007) reviewed the research findings from the past twenty-five years on end-user searching behavior. The author concluded that system designers need to utilize research findings when building systems “that are sensitive to the progress users are making in their ongoing searches, intervene with complex search features that are likely to solve user problems, and monitor users to determine whether these complex features help them achieve their goals.” The author cautions that it is important for Information Retrieval (IR) system developers maintain a level of simplicity with online IR system interfaces as they seek advancements in the searching capabilities of these systems.

At the 2003, Human Factors in Computing Systems Conference, several researchers and user interface designers presented a list of best practices that are summarized by Resnick and Vaughan. The search design best practices are divided into five categories, they are:

- **Structure of the Database**
- **Matching Algorithms**
- **User Content and Task Requirement**
- **Interface between the Information Seeker and the Search System**
- **Emergence of Hardware and Bandwidth Challenges with Mobile Devices**

These best practices were discussed in detail in the reviewed literature above.

The **sixth category** examined the **future role of catalogers and indexers**. Participants were asked whether or not they believe that the future role of catalogers and indexers is minimized in full text and metadata search systems.

A summary of the future role of catalogers and indexers in full-text search systems with metadata, include:

- Catalogers are still needed for descriptive metadata
- Human catalogers will play a lesser role in subject cataloging
- Catalogers needed to improve descriptors and identifiers for effective research and retrieval
- Catalogers view their work as art form and are not necessarily connecting people to information

- Catalogers and indexers working together with developers could make a better system

A summary of the future role of catalogers and indexers in metadata search systems include:

- Indexers are needed to add terms later for new concepts and changes in author names
- Always a need for human review to ensure accuracy
- Indexers still needed to identify terms excluded from the text
- Metadata indexing does not improve search results except for bibliographic metadata such as date, title, and author

A summary of the reviewed literature below, discusses the role of catalogers and indexers in the future, and also the overall role of online catalogs in light of other discovery tools.

Borgman (1996) noted that information seekers still find online catalogs difficult to use, even though there has been improvements in interfaces. One fundamental problem with these catalogs is that they were designed for the skilled or experienced intermediaries, and not the end users. Another observation made by the author is that studies on “information seeking” have demonstrated that searchers or information seekers conduct their query in stages. First questions are formulated in stages that are articulated in a query. A search may be conducted over a number of sessions using different information technologies and sources, both online and offline, with multiple options to answer a question or address an issue. The designs of most online catalogs are based on the assumption that information seekers formulate a query that represents a fixed goal for their search, and that each search is independent of the other.

Another dimension to the online catalog debate was posed by Deanna Marcum in 2006, when she asked the question, “do we need to provide detailed cataloging information for digitized materials, or can Google be viewed as a catalog?” The high cost of cataloging and its shrinking use may dictate its future.

Calhoun in 2006, suggest that while there are “prevailing strategies to integrate the catalog with other discovery tools, there is some reluctance from research library leaders, their staff, and university faculty members to such a change.” There are initiatives by Google, RedLightGreen, and Open WorldCat to expose research library collections on the web. Federal agencies have also adopted similar approaches in order to increase the visibility of their collections and to improve open access.

The **seventh category** examined the **role of metadata in improving search results**. Participants were asked several questions related to this issue.

A summary of the role of metadata in improving search results include:

- Improved availability of metadata provides better search results
- Provides the structure necessary to allow for consistency in quality searches

- It will improve the accuracy of search results
- It offers fielded search, by organizing search results into categories and displaying consistent search results.
- It is one of many sources of evidence for searchers that is too expensive to produce manually
- Structure controlled vocabulary to get to concepts
- Cross-referencing
- It helps in narrowing and refining search results
- Metadata should automatically map to content

A summary of response as to whether full text searching eliminate the need for metadata include:

- Full text searching makes it possible to search without metadata
- It is an opportunity to enrich the text
- Metadata may often have other non-searching roles to play
- Metadata is the way to add information that does not appear or is not readily discernable from the document itself
- Full text searching demands adding meta tagging to make the searching useful
- Metadata is still needed for classification and limitation on the document

A summary of responses as to whether full text searching can be used to effectively augment metadata include:

- A mix may be the optimal approach
- The two are often used together successfully to produce more precise search results
- They should complement each other

The final issue that was addressed in this study was an investigation into searching in the future. What will web search in the future provide that is not currently available? What do researchers think search technology will provide as advancements in search systems and information access improves?

There are differing views on searching in the future. Both technological advancements in search systems and improvements in information harvesting across multiple databases on a global platform will play a major role.

DuPuis (2006) suggest that the future can be viewed from two approaches. The first approach is to view things from “how we think things are going to be.” The second approach is to look at things from “how we would like things to turn out.” The author believe that future information seekers “net generation,” and beyond will not have the level of attachment to journals, conferences, and monographs, instead, they will have “expectations of simplicity.” There desire is to find rather than to search. Both publishers and database providers are now beginning to accept the fact that information seekers do not care where the information resides, they merely want to find the information that is needed.

Future searching must address the question of how to improve information access. Interactive and visualization tools will probably play a paramount role in demonstrating relationships among entities in multi-dimensional forms. Expectations should include improved search engine interfaces that will enhance and deliver seamless results. Scholars have suggested that human interaction with machines will improve so that future search engines will be able to understand the information seeker behavior pattern regarding searching, and anticipate the expected search results.

It is the hope that improvements to human-computer interactions and a comprehensive assessment of the ambiguity of images, words and objects should enable unified access to information across multiple platforms.

With better understanding of information seeker behavior, more effective search engine interfaces can be developed that will lead to improved search results. One constant has been the information seeker search strategy on the web. Not much has changed except for their unwillingness to view more than a page of search results.

McKay believes that future searching will be universal, pervasive and necessary. The author expects technical boundaries to disappear and searching will be available “everywhere all the time. It will be universal.” Consumers and information providers (both government and business) will have more access to information about individuals. Peterson (2005), also supports this view, by noting that search engines are now building profiles on information seekers. The net effect is that companies are now getting access to the collective wants and needs of the population. Consumer demand will increase with greater expectations.

In Rose and Levinson’s (2004) study on the “understanding of user goals in web search,” the authors suggested that by capturing information on the user behavior pattern, such knowledge can be used to modify search engines algorithms and interfaces that will lead to improve search results.

The choice of access tools to information on the web has moved to cellular phones, mobile devices, and the television, however, limitations in bandwidth remains an unresolved issue. Researchers predict that some day, television and searching will merge. This will allow simultaneous access to broadcast programs and the capability to search for additional information as needed. Current access to video within a search is a step in this direction.

Sokullu (2006) believes that internet searching is still in its infancy, as attempts are being made to find better searching and indexing techniques. The author sees **three trend areas** in the search industry, they are: *user interface (UI) enhancements; technology enhancements; and approach enhancements (vertical search engines)*.

Bourdoncle (2007) noted that among the issues and challenges that we currently face, is how to improve consumer user interfaces. They are viewed as too simplistic (search/browse result list/next page), only good for unstructured web pages. He calls for a universal browsing tool for semi-structured information.

Mostafa (2005) noted that search engines have mastered two major hurdles in information retrieval; their ability to handle large scale web crawling tasks; and, indexing and weighting methods. This has resulted in improved ranking results. The author believes that online search engines will soon provide major enhancements that will change how we find what we need.

Mostafa predicts that the next generation search technologies will include more powerful tools that combine search functions with data mining operations that will be able to look for trends or anomalies in databases without actually knowing the meaning of the data. Information seekers will be able to search through multiple data repositories by using visually rich interfaces that focus on broad patterns on information rather than picking out individual records.

In summary, the review of current literature on search methodology and search systems with the responses from interviews conducted with 48 participants across 29 organizations and institutions highlighted some thought provoking views as to where the future lies in search technology. There are still barriers to overcome to improve the information seekers' search experience on a global level with seamless access to multi-dimensional data.

APPENDIX: A: Table Summary

SEARCH METHODOLOGY STUDY

PREFERRED METHOD OF SEARCHING

Question # 17 & 18

CENDI MEMBER AGENCIES RESPONSES

Table # 01

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS # 17 & 18. Preferred method of searching and rational
Defense Technical Information Center, (DTIC)	A				X	Don't have one!
Defense Technical Information Center, (DTIC)	B			X		It depends on what I am looking for! For example, if I am looking for a specific fact, full text searching may be the only way to find it. For example, looking for DOD technical reports that are not in our collection, usually the only way to do it is to look on a general web search engine and look for organization, since that is one of the few search types they can do. Looking for specific known documents, I would use author or title metadata. Looking for general information I might use either full text or metadata. I lean more towards metadata searching for search engines that have that capability.

CENDI MEMBER AGENCIES RESPONSES

Table # 01 (Continued)

Defense Technical Information Center, (DTIC)	C	X				Because metadata is incomplete and inaccurate.
Defense Technical Information Center, (DTIC)	D	X				The fear that the search terms that I use may not be part of the controlled vocabulary, therefore my results will be minimized or fruitless. Metadata is not cost effective. The results may reflect poor cataloging. The recall for full text searching will always be better than that of metadata searching. What users really need is both full text searching with metadata tags.
Defense Technical Information Center, (DTIC)	E		X			If data is input correctly, then results will be more relevant in a metadata search system than with full text searching.
Defense Technical Information Center, (DTIC)	F	X	X			With metadata searching, one gets better results! Do not have much opportunity to do full text searching in my current job!

CENDI MEMBER AGENCIES RESPONSES

Table # 01 (Continued)

Defense Technical Information Center, DTIC)	G			X		A combination of Boolean fielded searching including controlled vocabulary, Boolean full-text searching, and algorithmic full-text searching. Proquest is an example of this.
Government Printing Office (GPO)	A			X		It depends on the information source and my knowledge of the information being searched for as to which I would choose to use.
Government Printing Office (GPO)	B	X	X			Dependent upon how the information is tagged and presented, the user is afforded a better opportunity of achieving relevancy in their search query.
Library of Congress	A	X				Get a lot of ideas! Helps me structure, gives me ideas!
Library of Congress	B	X	X			Depends upon search, but normally I'll do a full text keyword search; once I have found a relevant article I'll use the subject field to find more relevant articles.
Library of Congress	C	X	X			My method depends on the database and what I am searching for. I generally cast a wide net at first and narrow my focus as I gain a better understanding of what I am looking for and what is available. My preferred 'wide-net' tool is Google. Beyond that it depends completely on the subject being searched.
NASA Scientific and Technical Information Program	A			X		Metadata or full-text, depending upon requirements, time, and what I am looking for...also highly dependent upon my requirement, what I am searching in and with, and a number of other factors.

CENDI MEMBER AGENCIES RESPONSES

Table # 01 (Continued)

National Agricultural Library (NAL)	A			X	I want it all in one, as with science.gov Comprehensiveness of coverage, ease of use
National Archives and Records Administration (NARA)	A	X	X		Benefits of both full-text and metadata searching. Ideally, a system would allow for both. Prefer full-text keyword searching at the beginning search session. May do an initial survey and see how many hits I can retrieve. I typically narrow my search by using Boolean operators, wildcard characters, or nesting to account for variations of a keyword in the full-text searching environment. Once I have narrowed my results set, I browse through the results and make judgments about the relevancy of the hits. If the metadata or indexing is displayed as part of each result, I would take notice of the controlled vocabulary terms that appear in the results that interest me the most and most clearly match the topic of my research. I might try to search again with those terms.
National Archives and Records Administration (NARA)	B			X	I think it depends on the database and the quality of the data and/or metadata. Sometimes a full text search is effective, and sometimes it is not enough.
National Library of Medicine (NLM) NIH	A	X	X		It depends on the kind of information, but I will generally choose a combination of metadata and full text searching, with the metadata more heavily weighted than the full text.
Office of Scientific and Technical Information (OSTI) DOE	A			X	It would depend on the topic or area to be searched. Full Text Search capability is immensely helpful when looking for a needle in a haystack – when the classification or structure or hierarchy is not known but a small amount of very precise information is available

CENDI MEMBER AGENCIES RESPONSES

Table # 01 (Continued)

Office of Scientific and Technical Information (OSTI) DOE	B	X				I like the widest starting point possible.
USGS Biological Resources Division (Dept. of Interior)	A			X		It depends as to what I'm looking for. If I'm looking for a multi-media item, I want to go through the metadata – I do not have time to wait on the large files to open/download/view. If I'm looking for a publication, my initial preference is to search the metadata in that I believe, if it is categorized properly; I will get more targeted results. If I do not understand what is in a repository, I would prefer to do a full-text search to at least get some initial results to understand the content, structure, and organization. I may then go back and narrow through using the metadata.

SEARCH METHODOLOGY STUDY

PREFERRED METHOD OF SEARCHING

Question # 17 & 18

DOD Organizations and DOD Contractors

Table # 02

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 17 & 18. Preferred method of searching and rational.
Air Force Research Laboratory WPAFB	A		X			It will be more precise
Chemical and Biological Information Analysis Center (CBIAC)	A			X		It would depend on the topic or area to be searched. Full Text Search capability is immensely helpful when looking for a needle in a haystack – when the classification or structure or hierarchy is not known but a small amount of very precise information is available
Chemical and Biological Information Analysis Center (CBIAC)	B.	X				I like the widest starting point possible.

DOD Organizations and DOD Contractors

Table # 02 (Continued)

Johns Hopkins University, Applied Physics Laboratory	A		X		X	Better Results!
Johns Hopkins University, Applied Physics Laboratory	B		X			Response time! More precession!
Johns Hopkins University, Applied Physics Laboratory	C			X		More Precision!
Lackland Air Force Base	A				X	As long as I understand the way a search engine works, I can use any database and feel that I am effective with the search results I receive.
MITRE Corporation	A	X				Good metadata is not widely supplied with government information and searching by metadata requires you to know the appropriate government jargon to match.

DOD Organizations and DOD Contractors

Table # 02 (Continued)

MITRE Corporation	B	X	X			Use full text as a vetting process, i.e., Google Scholar; next apply metadata searching to improve search results!
Naval Research Laboratory (NRL)	A			X		I seem to get the best results when the 2 methods are combined.
Pentagon Library	A	X				Easier to search!
US Army Library Picatinny Arsenal, NJ	A	X	X			Most databases I search use metadata/bibliographic info for indexing. I am more used to this type search. Full text searching always results in too many hits.
Redstone Scientific Information Center (RSIC)	A				X	No Comment Received!
Redstone Scientific Information Center (RSIC)	B	X				No Comment Received!

SEARCH METHODOLOGY STUDY

PREFERRED METHOD OF SEARCHING

Question # 17 & 18

UNIVERSITY PROFESSORS RESPONSES

Table # 03

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS # 17 & 18. Preferred method of searching and rational
Old Dominion University	A	X	X			I prefer to use metadata based search when I want to search for example documents from a particular author. I use full text search when I am exploring and not sure about the author on the subject classification.
Old Dominion University	B				X	I use what is convenient and available. Few systems I use give me a choice. It is only as a researcher that I stop to ask myself how the system works.
Syracuse University	A		X			No comments provided!

UNIVERSITY PROFESSORS RESPONSES

Table # 03 (Continued)

Syracuse University	B	X	X			Prefer to have the choice of full-text and metadata-based search. Any one method alone will be inefficient.
San Jose University	A		X			Government information is very complex. Only metadata can best organize the information for retrieval.
University of North Carolina	A			X		It depends on what I'm looking for....for most US government websites (not databases), I am happy to navigate if the site is well organized (e.g., BLS has a lot of links on home page but they are well organized and explicitly stated and common data is a click away---quite browsable; others I'd search

SEARCH METHODOLOGY STUDY

PREFERRED METHOD OF SEARCHING

Question # 17 & 18

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 04

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS # 17 & 18. Preferred method of searching and rational
Access Innovation Inc.	A				X	Right now they are all essentially the same in presentation
Information International Associates Inc.	A	X	X			If you don't know the system then the easiest, fastest way is full text. The more one is a power user and understands the system the more effective metadata can become.
Information International Associates Inc.	B			X		I think a combination gives the best of both worlds and is the most likely to support both high precision requirements and high recall requirements. Which requirement is uppermost depends on the type of question the user is asking.
National Federation of Abstracting & Information Services (NFAIS)	A				X	Preferred methodology should depend upon the nature of the information required at any particular time. There will be times when full-text is absolutely required and other times when an amplified "abstract" or surrogate of the full text (i.e. One that contains an intimation of the conclusions reached in the research paper or a graphic that illustrates a particular region) will be adequate.

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 04 (Continued)

National Commission of Libraries and Information Science (NCLIS)	A	X				With new topics you need to have full text as it takes time for new terms into the controlled vocabulary.
Southeastern Library Network (SOLINET)	A				X	It depends on what I am researching for.

SEARCH METHODOLOGY STUDY

PREFERRED METHOD OF SEARCHING

Question # 17 & 18

OTHER LIBRARIES RESPONSES

Table # 05

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS # 17 & 18. Preferred method of searching and rational
Catholic University of America	A			X		I use item # if I have one, or sudoc. #
Senate Library	A		X			I can get more precise search results by using metadata.

SEARCH METHODOLOGY STUDY

SEARCHING METHODOLOGY STATUS

Question # 01 & 02

CENDI MEMBER AGENCIES RESPONSES

Table # 06

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 01 & 02 Searching Methodology...Full Text, Metadata, Other
Defense Technical Information Center, (DTIC)	A				X	<p># 1. There are a number of technologies that use various techniques to improve searching. They use complex algorithms and formulations, for example FAST. This search engine looks at word relationships. It looks at the frequency of words in the context it is used and then narrows down, or drill down. There are tools that address frequency counts, retrieved text, or most frequent words or phrases used. One would then click on a phrase to modify their search and then drill down. This is more effective than indexing and controlled vocabulary. Controlled vocabulary cannot cope in the changing environment. We tend to get false drops from DTIC TR.</p> <p>#2. Yes, by using algorithms and some metadata.</p>
Defense Technical Information Center, (DTIC)	B			X		<p>#1. But Google does not ...maybe three fields! URL and Title, can search separately in Google. Yahoo use to have everything in categories but switched to word searches because of all the manual labor. Too much recall with full text searching!</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 06 (Continued)

Defense Technical Information Center, (DTIC)	B (Cont.)			X		#2. Probably true. One kind of search will not do for all people. Note that the bibliographic databases started with metadata and only later turned to full text, mostly limited by technology at the time.
Defense Technical Information Center, (DTIC)	C	X				<p>#1. Full Text Searching has limited metadata. Controlled vocabulary is not uniformed across organizations leading to inconsistency. In the perfect world, controlled vocabulary would be universally applied and would provide optimum search experience. General Web content is not meta tagged. Full text searching allows information to be found, overcoming limitations with the application of controlled vocabulary. Extending the utility of full text search, vendors add relevancy methodologies beyond the content of the target document. Google adds a relevancy waiting based on a weighted value of sites that link to the document. A text crawler can categorize and add meta tags to documents based on the directories they are found, the sites they are located, the document type, date stamp, and the other documents in the collection.</p> <p>#2. The solution must address inconsistency in taxonomies. When there are no complete and universal taxonomies automatically generating meta tags at content creation, and where taxonomies exist they are inconsistently applied generating spotty search results. By mixing full text and meta tagging, search results can be improved, but the taxonomy must be consistent and consistently applied.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 06 (Continued)

Defense Technical Information Center, (DTIC)	D	X			<p>#1. I agree. For example, Google (Full Text search engine) relies on sophisticated relevance ranking algorithms to supplement their full text searching. Traditional full text searching thought of as a better fit for scientific data versus the social sciences and/or humanities.</p> <p>#2. I agree. Taxonomies can now be controlled, or system generated. Either way, they can be used to facilitate full text searching (e.g. categorizing search results into ‘buckets’.) Relevance ranking algorithms are probably the most important distinguisher (for me) on why I use one search engine versus another. Our users will desire that certain fields need to be field searchable. There are other fields that will probably be made redundant once we implement full text searching and will no longer be needed.</p>
Defense Technical Information Center, (DTIC)	E		X		<p>#1. Full-text databases don’t always incorporate legitimate structure, algorithms, etc. Maybe some do, but others, for personal or economic reasons, have developed algorithms that are almost considered trade secrets. Google is known for bringing up the most popular hits. Some full-text search engines, especially commercial search engines, tend to have current, suggested phrases to influence users as they search. Many commercial search engines show parallel results of related links for their advertisers and make pop-up items come up first with a searcher’s results.</p> <p>#2. That could be the case that we are approaching more of a blur in searching methodology. On the whole, most users don’t use Boolean. Instead, most users input a phrase or a word or two without checking the rules of the database. They are probably not aware of any algorithms.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 06 (Continued)

Defense Technical Information Center, (DTIC)	F	X	X		<p>#1. True Statement! Need some type of tool to get relevant information! With a true full text search engine, there is a tendency to get large number of false hits! Full text on its own, does not occur any more. Full text searching is good when one searching for a specific item, such as, a formula! We are now combining searching methodologies!</p>
Defense Technical Information Center, (DTIC)	G			X	<p>#1. True. Most full-text databases are no longer pure full-text devoid of structure but actually have metadata searching features. However, I understand that for the purposes of this discussion, the term “full-text” refers only to pure full-text databases, which are practically extinct. Even Google takes of advantage of XML and HTML tags and should not be used as an example of a pure full-text database.</p> <p>#2. Web programmers applied metadata and taxonomies because they encountered the same problems that library science encountered in the past, and applied the same solutions. An elusive special recipe of metadata, taxonomies, and algorithms is not going to generate 99% accuracy for searchers. Algorithms find approximate results that second guess the searcher’s intentions. They are good for exploratory searches to find at least something on a topic. Metadata and taxonomies might enhance this but they are also needed independent of algorithms for precise, controlled, and comprehensive searching.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 06 (Continued)

Government Printing Office (GPO)	A			X		<p>#1. A good XML structure is essential to fully unlocking the potential of the document regardless of the complexity of the search algorithms used. It clarifies the intent of the creator to an extent that will truly add value to the results.</p> <p>#2. There is certainly a large body of those who believe so. Applying commonly understood and interoperable indexing aides is the very essence of being able to increase the value of web search results.</p>
Government Printing Office (GPO)	B	X	X			<p>Each offers advantages and strengthens metadata for customers when combined. It is equally important that search include not just textual content of documents but metadata itself. Boolean methodology is antiquated and extremely limited in terms of meeting searching needs, regardless of the algorithm.</p>
Library of Congress	A	X				<p>#1. Agree! Most full text databases do incorporate some metadata.</p> <p># 2. At the Library of Congress we supply three subject headings for books, regardless of the size of the book that is not good enough! The Library is trying to figure out what strategy to applying for the various scenarios</p>
Library of Congress	B	X	X			<p>#1. Basically, when searching full text you are searching bibliographic fields such as title, subject... Depending upon the researcher's need and subject full text searching may suffice. However, searching controlled vocabulary is a more focused search, but this too has its limits.</p> <p>#2. Searching full text and bibliographic citations seems to be the way to go. There are multiple ways to search. Again it all depends upon the needs of the searcher. There is not a "one fits all" model in terms of information seeking and retrieval. But the more options a searcher has to search, the better.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 06 (Continued)

Library of Congress	C	X	X		<p>#1. I am not sure that ‘scholars’ think much about controlled vocabularies or any of the other details underlying search systems, except for scholars of information science and systems. The scholars I work with just want to know that they have found everything they need for their topic, and if they feel unsure in their search approach, will quickly ask a librarian for assistance. The scholar/researcher is rarely interested in the nuts and bolts of the databases they use, but the good ones are always confident that librarians know where and how to look.</p> <p>#2. I don’t think any of these elements are mutually exclusive. I do agree that the structure of a metadata system is very important, and especially so in a field that already has a unique vocabulary, for example the NGDC standard for geospatial metadata. The problem is enforcing the data input quality, which is the same issue that the Library of Congress has been addressing for forever with subject cataloging and MARC standards for data input. Or DTIC and its descriptors, etc. I think most studies show that the most effective systems are a combination of full-text and controlled index searching.</p>
NASA Scientific and Technical Information Program	A			X	<p>#1. This is largely true, though the mix may vary greatly between database applications. Not sure if most “scholars” would know much of this or explain in the terms above.</p> <p>#2. I think the real issue at this time, and given current technologies, is optimizing the recipe or mix of metadata, etc.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 06 (Continued)

National Agricultural Library (NAL)	A			X	<p>#1. I agree that people do not always understand the work that is being done in some full text searching applications to structure the text for better retrieval, and suspect that clearer and better information about the applications would improve searchers' understanding (for people interested in understanding).</p> <p>#2. I believe that the issue is not which recipe to apply, but rather how to present search choices most simply.</p>
National Archives and Records Administration (NARA)	B	X	X		<p>#1. Many real world databases, such as library and archival catalogs, make use of full text searching in their keyword search feature. Many less experienced (or impatient) researchers prefer a full-text keyword search because of their familiarity with Google's keyword search function. More experienced researchers and staff (librarians or archivists) might prefer searching based on specific bibliographic fields and controlled vocabularies. Having an access system that can accommodate both preferences is important for many organizations.</p> <p>#2. In database environments like bibliographic catalogs, full-text journal databases, and other "deep Web" databases not searchable via web search engines, a combination of full-text and metadata searching is prevalent.</p>
National Archives and Records Administration (NARA)	C	X	X		<p>Do not agree that a full-text search takes advantage of classification fields, abstracts, etc. A full-text search for "Jackie Kennedy" will not bring back a catalog record that has a subject access point for "Onassis, Jacqueline" (assuming the words "Jackie Kennedy" does not appear elsewhere in the catalog record.) However, if the search was a controlled vocabulary search for "Jackie Kennedy" (and this was a 400/variant name for "Onassis, Jacqueline"), it would bring back the catalog record.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 06 (Continued)

National Library of Medicine (NLM) NIH	A	X	X			No Comment!
Office of Scientific and Technical Information (OSTI) DOE	B				X	<p>#1. I think we have continued to evolve because of technology which has allowed us to reduce the investment in controlled vocabulary, taxonomies, subject classification, metadata, etc.,. XML is dominating the landscape as the markup language of choice. It is what goes on behind the scenes in the technology of building the database that is dominating and is changing the input requirements for putting the data/information in the database</p> <p>#2. That recipe also depends on the data/information types that comprise the database. Therefore the one-size fits all really does not apply on the technology side, however the average user does not know this. They want to perform searches easily and in a consistent and understandable manner.</p>
Office of Scientific and Technical Information (OSTI) DOE	C	X				<p>#1. Whether or not a database includes all the wonderful things that librarians traditionally relied upon, it is still possible to use complex search algorithms.</p> <p>Everyone agrees that better searching is enabled by richer databases. While bringing the collection under bibliographic control would further improve search capabilities, no one can afford to do it.</p> <p>#2. Searchers are well advised to take advantage of whatever recipe that the databases being searched will support.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 06 (Continued)

USGS Biological Resources Division (Dept. of Interior)	A			X		<p>#1. I believe full-text searching is really referring to indexing/searching the full-text/content of a document (word, pdf, html, etc.).</p> <p>I think the real issue may just be the amount of data/volume of data/information we have to deal with from a user/retrieval point of view. We are presented with so much data and information through search results that it is hard to distinguish as to what the best (highest quality), authoritative, and specific item we are looking for.</p> <p>#2. We use metadata to try and parse out results to users in different views, so that they are not necessarily overwhelmed with the search results.</p>
---	---	--	--	---	--	--

SEARCH METHODOLOGY STUDY

SEARCHING METHODOLOGY STATUS

Question # 1 & 2

DOD Organizations and DOD Contractors

Table # 07

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 01 & 02 Searching Methodology...Full Text, Metadata, Other
Air Force Research Laboratory WPAFB	A		X			<p>#01. I think that's true. Please consult this book – Ambient findability / Peter Morville – review at http://www.istl.org/06-summer/review4.html</p> <p>#02. There were sophisticated search engines before there was a web. DTIC, Dialog, and many others pioneered in Boolean search.</p>
Chemical and Biological Information Analysis Center (CBIAC)	A			X		<p>#01. Full Text Search capability is immensely helpful when looking for a needle in a haystack - when the classification or structure or hierarchy is not known but a small amount of very precise information is available - as with the source of a quote. Search algorithm as applied to full text searching is very different from the kind of hierarchical, taxonomic, classification-based approach one takes when reviewing the literature for a specific topic.</p> <p>#02. Yes, with the goal of keeping the widest range of approaches available to provide flexibility for the user.</p>

DOD Organizations and DOD Contractors

Table # 07 (Continued)

Chemical and Biological Information Analysis Center (CBIAC)	B.	X				<p>#01. Agree!</p> <p>#02. No Comment!</p>
Johns Hopkins University, Applied Physics Laboratory	A		X			<p>#01. Most full text databases do not have a good controlled vocabulary! Do have metadata, but not necessarily controlled vocabulary.</p> <p>#02. A knowledgeable searcher will find the information that they are seeking. It is important to have an underling structure with taxonomy and algorithms.</p>
Johns Hopkins University, Applied Physics Laboratory	B		X			<p>#01. Totally devoid, but I do like to use a controlled vocabulary. I don't believe that full text databases incorporate classification or structure.</p> <p>#02. Agree! I find value in the taxonomies!</p>
Johns Hopkins University, Applied Physics Laboratory	C			X		<p>#01. Don't know! Don't search enough! Create databases!</p> <p>#02. Algorithms stink! Yes! A recipe plus abstract and taxonomy.</p>

DOD Organizations and DOD Contractors

Table # 07 (Continued)

Lackland Air Force Base	A				X	<p>#01. Full text searching is a little more Google-esque than the research methods taught in Library Schools. The attempt with full text searching is to make searching through huge databases and the Internet an easier task for novices, and those not trained in the idiosyncrasies of particular database search engines operating features and functions.</p> <p>#02. Some sort of algorithm must exist for full text searching to accomplish its tasks for the searcher. Both methods (full text vs. metadata searches) contain an organized and controlled type of search method in order to return results. Yes, the real issue is what mix of metadata, taxonomies, and algorithms to apply.</p>
MITRE Corporation	A	X				<p>#1. Most search engines do look and weigh such fields as title, if they are supplied as a metadata tag, and others can be added to the calculation of relevance as appropriate.</p> <p>#2. I believe that there is no “one size fits all” solution. In some environments, the use of metadata and taxonomies may be appropriate; in others, such a fixed structure is not appropriate because of the time and effort required to establish, evolve, and maintain taxonomies. In addition, such taxonomic structures tend to be frozen in time and thus antithetical to discovery; they tend to be one person or group’s view of the information’s organization; and they are generally implemented with little understanding of the end users’ information seeking behaviors.</p>
MITRE Corporation	B	X	X			<p>#1. Web-search does not advance search...need state of practice. From a research perspective need to manage! From a practical perspective, need to promote capability. not always low arching! If capabilities are there, then they can be exploited. Few organizations exploit both full text and metadata searching capabilities. In the legal, genetics, technical fields, it is important to have both searching approaches.</p> <p>#2. The need to exploit searches, content extraction...semantics, hopefully</p>

DOD Organizations and DOD Contractors

Table # 07 (Continued)

MITRE Corporation	B (Cont.)	X	X			includes searches. I have not seen broad base semantics. Do agree searches need subject indexing. With better and more complex algorithms, though expensive, one will get better data extraction.
Naval Research Laboratory (NRL)	A			X		<p>#1. As search engines evolve, there is much better control and more appropriate retrieval generated. With the combination of bibliographic and full-text data, we can achieve increasingly better search results.</p> <p>#2. The combination of metadata, taxonomies and algorithm may vary in terms of the subject matter to be searched. In Sci/Tech searching, taxonomies and algorithm may be predominant keys in the strategy. In social sciences searches, the metadata elements may be most important followed by the taxonomies or algorithm.</p>
Pentagon Library	A	X				<p>#1. They do have controlled vocabulary hidden! Try using multi-searching. VISIMO and TEQMA use clustering!</p> <p>#2. Agree! Bibliographic and full-text combined! Agree with identifying the recipe of metadata to apply. Check article about LC on the Future of Cataloguing.</p>
US Army Library Picatinny Arsenal, NJ	A	X	X			<p>#1. I believe this to be true.</p> <p>#2. If both the full text and metadata are used for retrieval, then I think there needs to be some method of limiting search results; i.e. by author, data, or title. I don't want to get all the reports that list an author who is frequently cited in the references, I want reports <u>by</u> the author.</p>

DOD Organizations and DOD Contractors

Table # 07 (Continued)

Redstone Scientific Information Center (RSIC)	A				X	No Comment Provided!
Redstone Scientific Information Center (RSIC)	B	X				Searching full text seems to me to be the most effective if getting the correct information to the end user. OCR with full text search capability.

SEARCH METHODOLOGY STUDY

SEARCHING METHODOLOGY STATUS

Question # 01 & 02

UNIVERSITY PROFESSORS RESPONSES

Table # 08

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 01 & 02 Searching Methodology...Full Text, Metadata, Other
Old Dominion University	A	X	X			<p># 1. I am not sure what kind of full-text databases you are referring to. For me Google is one example of a very successful full text search database and it can not distinguish keyword search for different metadata fields like author, subject, abstract, etc. For example, if I search for “John” in Google, and if John appears in an abstract of a document, it will be consider a hit. However, user may wish to only get hits where “John” appears as one of the authors.</p> <p>#2. I agree partially. I still sometime prefer to use Boolean methodology in search (like use of phrase search in Google, which is an example of Boolean methodology).</p>

UNIVERSITY PROFESSORS RESPONSES

Table # 08 (Continued)

Old Dominion University	B			X	<p># 1. You offer a premise that you purport is a fact (most full-text databases incorporate...) but whose truth I am not at all convinced of, then use that premise to draw a conclusion that seems completely unrelated to that premise (therefore the definition of the phrase “full-text searching” is wrong). Even if your premise holds, the most you could conclude is that: either most searches ignore available information in the databases or that most searches of those databases are not, by definition, full text searches.</p> <p>#2. Again, the premise seems at best marginally related to the conclusion. In what area of computing would it not be critical to consider the question of what combination data, data structure, and algorithms to be employed? Given that the best “recipe” might very well involve ignoring any or all of the specific items you mentioned, the conclusion is nearly vacuous.</p>
Syracuse University	A		X		<p># 1. The real issue is what algorithms to apply to a given corpus for a given user community. Boolean operands only work well when controlled metadata is available. You can AND and OR to your heart’s content, but unless you are AND’ing on controlled terms, your effectiveness is going to be limited. The better approach is to know the community that will be searching the collection, know how to build an interface to let that community specify what they are looking for, and then build in algorithms that fill in search limiters previously found useful. The interface is the issue. The simpler the interface, the less knowledgeable the user, the more work has to be done by the search algorithms.</p>

UNIVERSITY PROFESSORS RESPONSES

Table # 08 (Continued)

Syracuse University	B	X	X			<p># 1. As a user, I do not care what controlled vocabulary a system uses as long as I get what I need. I think the users are taken for granted that the search system will take care of the problems to make the system work.</p> <p>#2. For domains that terminology is highly specialized, the role of metadata and taxonomies will be more critical since general purpose dictionaries will not be able to meet the specialized needs. Such needs are not only in the searching business, but also in the document categorization and representation process. That is, automatic metadata generation and indexing is the way to go; without taxonomies or other semantic knowledge bases it would be difficult, if not impossible, to automatically process documents and offer superior search performance. It is true that full-text search increased the flexibility and ease of search, but it also brought large amount of information noise. Using taxonomies and metadata can be viewed as a way to reduce the noise in search.</p>
San Jose University	A		X			<p># 1. We are talking about two different things: full text searching vs. full-text databases. While a full text database may incorporate some form of classification, its search engine may not always allow one to search that way.</p> <p>#2. Sorry, not sure about the question.</p>

UNIVERSITY PROFESSORS RESPONSES

Table # 08 (Continued)

University of North Carolina	A			X		<p># 1. Google et al have pretty much demonstrated that page rank or probability ranks are better than controlled vocabulary and metadata--- metadata is great for faceted search from a database-driven corpus like an e-commerce site.</p> <p>#2. Of course, as the search engines say---it is the secret sauce that differentiates search services</p>
------------------------------	---	--	--	---	--	---

SEARCH METHODOLOGY STUDY
SEARCHING METHODOLOGY STATUS
 Question # 01 & 02
INFORMATION SCIENCE ORGANIZATIONS RESPONSES
 Table # 09

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 01 & 02 Searching Methodology...Full Text, Metadata, Other
Access Innovation Inc.	A				X	<p>#1. It has been repeatedly proven that text blobs without control give imperfect results in search. The better the structure applied to data the more likely a search will turn up relevant material. As the library and information science fields have increasingly turned to full text without control the other sectors of the economy and especially computer science, archives and web portal creation community have turned increasing control using taxonomies etc. The LIS group cries that they are reinventing the wheel – that we have done this for ages. While LIS are turning away from our tried true and proven methods – others are finding the same techniques on their own and adopting them.</p> <p>#2. Yes – Boolean works best with controlled vocabularies. It is a two level activity – apply control when the materials are ingested. Use that control in the Boolean search. Gives excellent precision and recall the two conflicting ends of relevance.</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 09 (Continued)

Information International Associates Inc.	A	X	X		<p>#1. I agree that technology and semantic and syntactic models used in conjunction with taxonomies, ontology's, etc are getting increasingly sophisticated and able to gain more recall and precision in searching. When it is better than full text depends on what the questions and context are, but the point that new technologies are not void of knowledge structure is correct.</p> <p>#2. Yes, it is recipe of the mix but there is also need for and existence of continuing advances in the underlying models on how to apply them.</p>
Information International Associates Inc.	B			X	<p>#1. I agree. It does, however, depend on the search engine. Google still does not use metadata to the extent that Yahoo does. Sometimes it depends on the context. For example, I think that many commercial search engines, like Google, Yahoo and MSN, are geared toward the popular Web, and, therefore, they aren't as successful in marrying semantic support to full text searching as they are when dealing with entertainment and news. I agree that most full-text databases incorporate some form of classification and structure, because of the very nature of authoring a document. There is a title at least, but again, the question is what difference does it make to a particular search engine, and can the user figure out how to make the most of it.</p> <p>I also wonder what the impact will be of new modes of communication like blogs and wikis. Do these modes change what we mean by full text and, therefore, redefine success again?</p> <p>Institutional repositories and e-print repositories have obviously made a big difference here. In general, I think they have done a better job of imposing bibliographic control via metadata and not so much controlled vocabulary. (continued next page.)</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 09 (Continued)

Information International Associates Inc.	B (Cont.)			X		<p>If you consider a grouping of publications on the web to be a full text database, then the degree to which metadata is applied varies greatly. I can't tell you where this came from, but I have in mind that less than 1% of web sites have any metatags, and many of these web sites could be considered full text documents.</p> <p>#2. Yes, of course, it is the mix of all these that are applied. The right mix isn't easy to achieve. In addition, it also depends on how well the metadata, taxonomies and algorithms meet the needs of an increasingly more diverse audience.</p>
National Federation of Abstracting & Information Services (NFAIS)	A				X	<p>#1. Search engines were explained to a mass audience as largely operating on the basis of pattern matching of text strings. It wasn't until Google emerged that we started seeing the massive Web audience introduced to the idea of special algorithms as a part of the search environment. Google search queries structured to search specific fields are infrequently used by the average individual so the perception lingers that it's JUST an instance of pattern matching.</p> <p>The other element of this is that users largely have no interest in knowing too many details of how the "black box" in any given technology works. They just want it to work for them without too much pain or effort.</p> <p>#2. Yes, with the understanding that there will be an on-going challenge for content providers to develop different recipes according to the needs of a specific community of practice. How users think about content drives how they will search for it. How they think about content should shape our interfaces, presentation of results and our platforms to better enable retrieval by these users.</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 09 (Continued)

National Commission of Libraries and Information Science (NCLIS)	A	X				<p>This is so true when you can put in your search elements that make use of the metadata such as domain name. Such as searching in Google when you limit your search to the 'gov' domain or 'pdf' file type.</p> <p>This could be for the general public as they sometimes look for a given format (CD of a book or the printed book).</p>
Southeastern Library Network (SOLINET)	A				X	<p>While I agree with the statement, the problem is the information itself and the way it is displayed. In most cases, it is just too much irrelevant information.</p> <p>Too many choices but limited training on searching methodology.</p>

SEARCH METHODOLOGY STUDY

SEARCHING METHODOLOGY STATUS

Question # 01 & 02

OTHER LIBRARIES RESPONSES

Table # 10

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 01 & 02 Searching Methodology...Full Text, Metadata, Other
Catholic University of America	A			X		# 01. Yes, some form, such as limiting to a certain field, which helps
Senate Library	A		X			#1. That's true in part, but unless users are aware of the controlled vocabulary terms used in the full-text database, they still are whistling in the dark. #2. Agreed. The problem is that people, including library administrators, want everything to work like Google—plug in terms and supposedly you're set. The work it takes to set up taxonomies and provide metadata tags is pretty staggering, especially if you are trying to do it retrospectively.

LIMITATIONS IN FULL TEXT & METADATA SEARCHING

Question # 22 & 23

CENDI MEMBER AGENCIES RESPONSES

Table # 11

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 22 & 23 Limitations in Full Text & Metadata Searching
Defense Technical Information Center, (DTIC)	A				X	# 22. Few drawbacks! It can give you more information than you desire! # 23. No drawbacks!
Defense Technical Information Center, (DTIC)	B				X	# 22. Not finding what you are looking for because of too many hits. Lack of synonyms! Not being able to differentiate, e.g. find Brown the author from the color brown! Specificity is lacking! Large number of hits, or false hits! For long articles, unless you break up the text into chunks, you may get hits with terms far apart in the article, not related to each other at all. # 23. Terms one is looking for may not appear in the citation. You can only use subject terms from the title, abstract, or controlled vocabulary. You wouldn't find a specific fact in a particular sentence (assuming it's in there). You sometimes find inconsistency in the controlled vocabulary assigned to a document, since it is done by humans. You may be looking on a document from a different view than the original author or indexer! Terms change over time.

CENDI MEMBER AGENCIES RESPONSES

Table # 11 (Continued)

Defense Technical Information Center, (DTIC)	B (Cont.)			X		You have to do the work somewhere for good retrieval. You can do it ahead by indexers or you can try to think up all the synonyms yourself upon searching. Or just browse through huge numbers of hits.
Defense Technical Information Center, (DTIC)	C	X				# 22. Lots of bad results with relevancy that is not meaningful. # 23. May not capture all results.
Defense Technical Information Center, (DTIC)	D	X				# 22. A frequent complaint about full text search systems is the large number of hits derived. However, poor precision is not that important as long as the relevance ranking is good. One issue is that searchers don't optimize their search strategy. # 23. To create the metadata is expensive. And even then, it is only as good as the quality and accuracy of the input. Controlled vocabulary is not widely used anymore. There are problems with metadata rules that users may not understand or have not become familiar with, which results in poor search output.
Defense Technical Information Center, (DTIC)	E		X			# 22. False results! Words searched or retrieved may not be relevant terms. For example; military terms or acronyms that are also actually common words are difficult to search. # 23. Terms may not be input correctly or consistently! Indexers may not be picking up the best terms for the documents. Misspelling when inputting data is also a problem since the input affects search results.

CENDI MEMBER AGENCIES RESPONSES

Table # 11 (Continued)

Defense Technical Information Center, (DTIC)	F	X	X		<p># 22. Relevance! Too many hits!</p> <p># 23. One may miss the most important document, since one is relying on the work of the cataloger and indexer.</p>
Defense Technical Information Center, (DTIC)	G			X	<p># 22. If it is “pure” full text searching (i.e. no option to limit your search to particular areas of the document and the algorithm does not take into consideration where terms appear in the document) then you will typically receive many nonrelevant documents. This is why nearly all full-text databases are hybrids that include fields or recognize areas of the document. If it is an algorithmic search of a full-text database, the disadvantage is lack of precision and control. Its purpose is only to find a few good documents. In both scenarios the searcher must guess all the possible terms people might have used to describe a topic in order to run a comprehensive or precise search. If controlled vocabulary had been used, the searcher would only have to look up and search the one term used by catalogers. Full-text searching also doesn’t allow basic searches that nearly every one needs: search by author, search by publication date, search by title, and search by type of document. Full-text searching is completely at the mercy of the author (or scanning software) and errors that they made. In metadata databases, catalogers often overcoming this by correct misspellings in titles and authors’ names and investigating authors’ pseudonyms.</p> <p># 23. Metadata databases do not allow you to find quotes nor every mention of a word or phrase in the full-text. They also are at the mercy of the people who enter the data, catalog the items, and assign controlled vocabulary terms. Institutions have control over this variable through training and quality control, whereas they usually have no control over the quality of the full-text.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 11 (Continued)

Government Printing Office (GPO)	A			X		<p># 22. It tends to be less accurate than well-constructed fielded searching.</p> <p># 23. It is useless if the user does not know the structure and meaning of the metadata.</p>
Government Printing Office (GPO)	B	X	X			<p># 22. I am not aware of any.</p> <p># 23. Lack of controlled vocabularies by the author.</p>
Library of Congress	A	X				<p>#22. Lead's to large irrelevant results.</p> <p># 23. Not always clear how system works! The way we describe terms!</p>
Library of Congress	B	X	X			<p>#22. High recall, low precision. May need to look at a lot of records before you find the relevant one! Normally full text searching does not yield many relevant results... after all you are searching for the occurrence of a word in a document- not what the document is about.</p> <p>#23. How good is your metadata? Is it minimal? Are you using controlled vocabulary? One relies upon the expertise of the human inputting the metadata...</p>
Library of Congress	C	X	X			<p>#22. Using the wrong word(s). Not knowing the right word(s). Irrelevancies, too much stuff, etc.</p> <p>#23. Not knowing the vocabulary or understanding the concept. Requires more education and thought.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 11 (Continued)

NASA Scientific and Technical Information Program	A			X		<p># 22. Results can become overwhelming, devoid of context and therefore less relevant, and not very time efficient.</p> <p># 23. You are totally dependent (and at the mercy) of whoever created the metadata, how well done is it, and does it accurately reflect the data it describes.</p>
National Agricultural Library (NAL)	A			X		<p># 22. Need relevance ranking, may need language translation facilities.</p> <p># 23. Results may not be as rich as with full text searching.</p>
National Archives and Records Administration (NARA)	A	X	X			<p># 22. Some of the drawbacks include a) end users' inexperience with full-text search strategies, such as the need to include variant forms and synonyms of a keyword, might lead them to feel dissatisfied with their search results; b) its relative imprecision compared to retrieval based on a controlled vocabulary indexed system (ex. false hits); and c) issues with relevancy ranking when a keyword does not appear frequently throughout a long text, but is a major topic.</p> <p># 23. A) End users' inexperience with metadata searching and b) less prominent topics that appear as part of the document or data but are excluded from the metadata, are examples of some of the drawbacks in using metadata searching.</p>
National Archives and Records Administration (NARA)	B	X	X			<p># 22. The "relevancy ranking" assigned can be inaccurate. A certain word or topic may be found via full-text searching and assigned a high relevancy ranking because it appears more often than another word/topic. But the document or record may actually be more about the latter topic. So full-text searching can be flawed or misleading.</p> <p># 23. The end user often has to be a more experienced searcher to do effective metadata searches.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 11 (Continued)

National Library of Medicine (NLM) NIH	A	X	X			<p># 22. Relevancy is often a problem with full text searching.</p> <p># 23. ... sometimes the searcher only has a small scrap of text to use in a search and no clue to the meaning of the document; in those cases, full text is the best way to search a document.</p> <p>If the person or program responsible for assigning metadata is not skilled, the metadata might be useless. Also, if the search function does not search both authority files and the descriptive metadata, searches might yield few to no hits (for example, if the user searches for books written by Samuel Clemens and the search does not match the author's name to his many pseudonyms, the results will not show all of his works). Metadata is also expensive to create, assign and maintain, so its quality varies greatly from database to database</p>
Office of Scientific and Technical Information (OSTI) DOE	A				X	<p># 22. I have not experienced any drawbacks to full-text searching.</p> <p># 23. Other than for the classes of documents or information I mentioned above I see little use for metadata in today's world. I was an early proponent of Dublin Core in 1995. As processing speed increased, budgets went down, storage got cheaper, and tools became more effective -- metadata for text documents increasingly fell away, in favor of full text searching using faster machines, cheap storage, and better index structures.</p>
Office of Scientific and Technical Information (OSTI) DOE	B	X				<p># 22. The biggest drawback is that you can't search by field.</p> <p># 23. The main drawback using metadata searching is that few can afford to create the metadata. For example, bringing the e-Print Network under bibliographic control would require more than OSTI's entire resources.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 11 (Continued)

USGS Biological Resources Division (Dept. of Interior)	A			X	<p># 22. To many search results, information overload for the user, missing results, improper classification of documents, slow response time, difficulty in presenting customized results to users, operation requirements (as with creating metadata – sometimes you might just be pushing the costs from human cataloging to hardware/software), and need to understand the content (as with creating metadata).</p> <p>Full-text searching is needed, but, as with metadata, it is not the answer for every information repository (need to fully understand the information content, delivery purpose, user needs, etc.).</p> <p># 23. Users may not understand the process used to classify a document requires significant up-front human resources, may require and additional user interface (vs. simple search box), may require training/examples, tips to aid the user, requires weighting of certain elements – to improve results, and typically a user has to understand the scope/intent/content of the repository better than with full-text searching.</p>
---	----------	--	--	----------	---

LIMITATIONS IN FULL TEXT & METADATA SEARCHING

Question # 22 & 23

DOD Organizations and DOD Contractors

Table # 12

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E	COMMENTS
						# 22 & 23 Limitations in Full Text & Metadata Searching
Air Force Research Laboratory WPAFB	A		X			# 22. You will retrieve irrelevant results – e.g. – a search on “xyz” will retrieve a result that says “this paper is not about xyz” # 23 Involves understanding the concept that a metadata search will retrieve the paper whether it uses the term “drone” or “remotely piloted vehicle”
Chemical and Biological Information Analysis Center (CBIAC)	A			X		# 22. Increases the chances of getting spurious results. Time consuming. #23. Lack of precision and flexibility is a possibility.
Chemical and Biological Information Analysis Center (CBIAC)	B	X				No Comment Received!

DOD Organizations and DOD Contractors

Table # 12 (Continued)

Johns Hopkins University, Applied Physics Laboratory	A		X			<p>#22. Having to “or” every way a word is used in order to get good results!</p> <p># 23. Controlled vocabulary is slow in updating! It takes a while for new terms to be accepted! Also, the use of author key words!</p>
Johns Hopkins University, Applied Physics Laboratory	B		X			<p># 22. High recall! Irrelevant material!</p> <p># 23. It may eliminate relevant data!</p>
Johns Hopkins University, Applied Physics Laboratory	C			X		<p># 22. Large number of hits!</p> <p># 23. Taxonomy may not be created well! It must support system for which it was developed.</p>
Lackland Air Force Base	A				X	<p># 22. Natural language idiosyncrasies, use of slang and jargon, abbreviations or acronyms that can have multiple meanings, misspellings.</p> <p># 23. Controlled vocabulary, unfamiliar with thesaurus</p>
MITRE Corporation	A	X	X			<p>#22. Poor precession and recall! Issues of awareness, performance and usability!</p> <p>#23. Inconsistency, expensive, time consuming, difficulty in keeping terms up to date in certain disciplines due to constant changes!</p>

DOD Organizations and DOD Contractors

Table # 12 (Continued)

MITRE Corporation	B	X				<p>#22. Search terms may not match jargon or business-specific terminology. Content may include a number of synonymous terms, depending on the author, where uniform use of terms would be better.</p> <p>Acronyms used in the content may not be familiar to users.</p> <p>Users may use a variety of ways to search for the same content, for example, on IRS.gov, we have identified more than a dozen search terms equivalent to the 1040-EZ form (e.g., ez1040, 1040ez, 1040 ez, form1040ez, e-z).</p> <p>mismatch of user terminology with jargon used in the content or with something that needs a fairly exact match, such as a form number.</p> <p>#23. See earlier answers</p>
Naval Research Laboratory (NRL)	A			X		<p># 22. As a single approach, it may not draw together the elements that will most quickly pinpoint a document</p> <p># 23. A relative little used term may not be included in metadata string if the metadata creator did not choose to include the term.</p>
Pentagon Library	A	X				<p># 22. Hit or miss terminology! Terms may only appear in the title which then results in a low relevancy! The ideal is to have both (full text and metadata) to complement each other, the best of both worlds!</p> <p># 23. Structured terms! Creation of terms! Whether the end user has the ability to relate to the subject terms used! More emphasis should be placed on the end users thought process!</p>

DOD Organizations and DOD Contractors

Table # 12 (Continued)

US Army Library Picatinny Arsenal, NJ	A	X	X			<p># 22. Getting too many hits because of citation listings.</p> <p># 23. Not being able to find documents on specific subtopic; i.e. M28 projectile info not found in a metadata search that indexed a document under projectiles.</p>
Redstone Scientific Information Center (RSIC)	A				X	No Comment Received!
Redstone Scientific Information Center (RSIC)	B	X				No Comment Received!

LIMITATIONS IN FULL TEXT & METADATA SEARCHING

Question # 22 & 23

UNIVERSITY PROFESSORS RESPONSES

Table # 13

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 22 & 23 Limitations in Full Text & Metadata Searching
Old Dominion University	A	X	X			# 22. Low precision # 23. Not user friendly
Old Dominion University	B				X	# 22. Low precision, low speed, unfriendly interfaces # 23. Low recall & precision, unfriendly interfaces, cost of acquiring accurate metadata
Syracuse University	A		X			# 22. No response provided # 23. No response provided

UNIVERSITY PROFESSORS RESPONSES

Table # 13 (Continued)

Syracuse University	B	X	X			<p># 22. Could be.</p> <p># 23. Too much information noise while possibly missing out important information.</p>
San Jose University	A		X			<p>#22. Precision rate is too low.</p> <p>#23. When everybody becomes information literate, metadata searching is not a problem.</p>
University of North Carolina	A			X		<p>#22. Too many hits, term ambiguity</p> <p>#23. Too few hits, missing categories, term ambiguity</p>

LIMITATIONS IN FULL TEXT & METADATA SEARCHING

Question # 22 & 23

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 14

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 22 & 23 Limitations in Full Text & Metadata Searching
Access Innovation Inc.	A				X	# 22. Getting results that have nothing to do with the users thoughts in the search query but are in fact accurate in the use of the terms used in the query. # 23. Expense of applying the metadata and allowing only the term deemed the preferred term in the search itself.
Information International Associates Inc.	A	X	X			# 22. Word control. Plant or plant of plant. Precision of search -- Fine tuning the search well enough to get what is really wanted. # 23. Human error in the construction of metadata (especially with controlled manually produced indexing)

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 14 (Continued)

Information International Associates Inc.	B			X		<p># 22. Pure full text searching is very dependent on the search engine. If you have a database that could be used with several engines, then you get more consistency if they use metadata. If you are looking for something very precise, but not a named entity, then you have to be a more skilled searcher. Those requirements are aided by the inclusion of metadata, especially if it has the power of a good thesaurus or ontology behind and takes advantage of it to produce things like synonym rings to expand the search.</p> <p># 23. Metadata searching can sometimes be too precise. In order to think of all the ways a user might approach the document, you need a real good indexer. Most indexing is done to be the most precise. This often makes it difficult to find broad concepts.</p>
National Federation of Abstracting & Information Services (NFAIS)	A				X	<p># 22. Language is the biggest drawback and consequently the volume of content retrieved.</p> <p>#23. Limited understanding on the part of the user as to what fields are included. Limited information that the user may have on hand with which to form the search query. Inconsistent data (some fields may not be populated).</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 14 (Continued)

National Commission of Libraries and Information Science (NCLIS)	A	X				<p># 22. Speed and use of system</p> <p># 23. May limit the task at hand.</p>
Southeastern Library Network (SOLINET)	A				X	<p># 22. The information layout. Researcher has to scan the full text to find the needed information.</p> <p># 23. Researcher must understand the way the information is presented.</p>

LIMITATIONS IN FULL TEXT & METADATA SEARCHING

Question # 22 & 23

OTHER LIBRARIES RESPONSES

Table # 15

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 22 & 23 Limitations in Full Text & Metadata Searching
Catholic University of America	A			X		# 22. Lack of precision. # 23. No Response!
Senate Library	A		X			#22. Lack of precision ... unless the words you are using are really precise themselves. See my comments about “automobiles” and “cars”. These kinds of searches pull up tons of false hits. #23. The taxonomy/thesaurus/metadata schema needs to be accessible to users or they won’t be aware of them; these tools also need to be pretty sophisticated (lots of references) or they won’t pull up arcane terms.

SEARCH SYSTEMS PERFORMANCE AND MEASUREMENT

Question # 09 & 10

CENDI MEMBER AGENCIES RESPONSES

Table # 16

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p># 09.... current state of search engines...</p> <p># 10. ... flaws in measuring search systems performance...</p>
Defense Technical Information Center, (DTIC)	A				X	<p># 09. Search engines are always making improvements to their algorithms. More metadata tagging, means better results. There are significant improvements in search results by improving the algorithms and by exploring more data.</p> <p># 10. It is hard to get test data sets of significant size in order to determine relevancy. One needs a large data test to get good results! There is also different rating and ranking among different search engines!</p>
Defense Technical Information Center, (DTIC)	B			X		<p># 09. ...Full text search engines need to go back to metadata to get more specific data, such as using author searching. It depends on what one is looking for. A user might well want to emphasize recall rather than precision.</p> <p>If search engines were already good enough, then why is each application using one different? Different uses and users for each one?</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 16 (Continued)

Defense Technical Information Center, (DTIC)	B (Cont.)			X	<p>Search interfaces need to be designed to allow different kinds of searches – retrieval of a specific document, all on a topic, a few good articles, specific fact. Note that it also depends on the collection content. For example, if there are no fact articles, then a fact search won’t get you anywhere.</p> <p>I agree that no search engine can be perfect. But they need to be flexible to allow different interfaces and capabilities for different needs.</p> <p># 10. Not sure how to measure! Precision and recall aren’t perfect. You do not know how to measure until you know all the hits that match the intent of the query. This is a huge job and limits the size of the search collection that can be measured in this way. I suppose you could have several people do the search and assume that the one that got the most hits is the best! There is no way to get perfect precision or perfect recall, though you could get perfect retrieval if you just retrieve the whole collection.</p>
Defense Technical Information Center, (DTIC)	C	X			<p># 09. Search systems are getting better but they are not good. Look up “DOD Blogs” in Google and you get almost 6 million hits. This is little way to limit the search in a way that would tell you if there is a list of DOD public blogs. So while all the 6 million hits might be accurate, they are not precise. It depends on how accuracy is defined. There is not one universal and valid relevancy ranking method.</p> <p># 10. Search engines are judged by their speed or results. This does not mean right results. There is no simple way to measure the quality of the result. A person may be presented with 100 results and may feel satisfied that they got useful information, however, might be unaware that they missed the most relevant and useful document. It is more a question of, do you find what you need. Looking for the relationship between two objects</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 16 (Continued)

Defense Technical Information Center, (DTIC)	C (Cont.)	X				might be critical, a full text search that is not tuned to connect the dots might miss the most important relationship or may bury it 10,000 hits down the list. If the search responds in less than 1 second, but then the user spends hours stepping through the results, the speed of the initial response is not meaningful
Defense Technical Information Center, (DTIC)	D	X				<p># 09. It's more than an accurate system. There are other important issues such as ease of use, recall, precision, intuitive interface, etc.</p> <p># 10. User evaluations are probably the most important measure. Traditional measures are recall and precision. Not sure if these measures have been updated to account for full text searching. For example, the classic definition of precision is no longer applicable – the important measure with respect to precision would measure how good the relevance ranking performed.</p>
Defense Technical Information Center, (DTIC)	E		X			<p># 09. I don't think that we will ever believe that our search systems are good enough! As our systems become more advanced, our expectations as users become higher. Our information retrieval systems are advanced enough, however, that one can skew the results to make the relevant items appear at the top of the results page according to a preferred algorithm.</p> <p># 10. One could question the purpose and the accuracy of the company or person who sets the algorithms for a given system. Also, one may question what factors are used in determining the relevancy ranking. In addition, some Web publishers purposely use incorrect metadata so that their information will be retrieved by searchers. We cannot overcome all of these issues since many search systems are motivated by economics.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 16 (Continued)

Defense Technical Information Center, (DTIC)	F	X	X		<p># 09. Search systems can always be improved! Users don't care about the search system as much as they do about quality of the interface.</p> <p># 10. Don't know how search engines measure their performance. Search engines tend to measure only their hits! You don't know if the user got hits! You only know that they got results! That is all the user status provides.</p>
Defense Technical Information Center, (DTIC)	G			X	<p># 09. Metadata databases tend to do, are supposed to do, exactly what you tell them with 100% accuracy. If I ask for all the reports with 'tank' in the title, it should retrieve ALL the reports with 'tank' in the title. The only exceptions should be due to errors in the data (which exist in both metadata databases and full-text databases) or errors in the indexing. You cannot get better than 100% accuracy but you can apply ranking to the results (as STINET and many other bibliographic databases offer). You can also improve the user interface to help the user to easily create better search statements. You might also improve the catalogers' application of the controlled vocabulary by giving them more time or training. Pure full-text databases, on the other hand, can try to improve their accuracy only by modifying their fuzzy algorithm. Once they incorporate metadata they become a hybrid with more options.</p> <p># 10. It is difficult to know the entire content of a database and whether or not searches are retrieving all the records/documents they are supposed to and whether or not one document should be ranked more relevant than another. If you are speaking of a typical internet search engine that uses an algorithm, the results could be compared to a straight Boolean metadata</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 16 (Continued)

Defense Technical Information Center, (DTIC)	G (Cont.)			X		search of the same content if that search capability is available. But then you are depending on the accuracy of the Boolean search and the quality of the metadata. Probably the biggest possible flaw in measuring search engine performance is using searcher satisfaction as a measure. Searcher satisfaction is a good measure of a search interface, not of retrieval.
Government Printing Office (GPO)	A			X		# 09. Just as a reference librarian can aide even the most experienced researcher, the refined nature of improved search will certainly assist in getting searchers to the right result. # 10. Users come in all shapes and sizes and even the best performance measurement of search can cover only a portion of the user universe.
Government Printing Office (GPO)	B	X	X			# 09. I disagree. I think that search engines are better than what they were, but more improvement is obviously needed, particularly for granular levels of content. # 10. Relying too heavily upon term/word appearance in metadata, particularly if this is coded into the HTML in terms of defining relevancy. Build an autonomous, intelligent agent that learns from both user actions and from the information content of queries and documents.
Library of Congress	A	X				#9. Searchers could yield more target results! For some searchers, the more targeted the search is the happier they are! # 10. Too many results from search! Presentation of results! Added value! Searchers can make own judgment from results, eg, Google!

CENDI MEMBER AGENCIES RESPONSES

Table # 16 (Continued)

Library of Congress	B	X	X		<p>#9. There is always room for improvement. However, search systems are good enough. If one takes the time to understand the searching fields and advance search features then searches will be more focused. I would like to see improvements on interface design and usability, making the search system more seamless.</p> <p>#10. Fundamental flaws in search engine performance. Searching only a selection of material (web search engines generally search only top level), relevancy ranking, minimal controlled vocabulary/ indexing (esp. in database searching), lack of multimedia searching within a document...</p>
Library of Congress	C	X	X		<p>#9 Sure. Some of them are good enough and some are completely inadequate. Sometimes design is really dependent on the nature of what is indexed and sometimes it doesn't matter as much. I think they all really need evaluation on a case-by-case basis. The characteristics of the evaluators have to be documented as well.</p> <p>#10. Not sure; the wrong evaluators? Targeting the evaluation to the proper user group.</p>
NASA Scientific and Technical Information Program	A			X	<p># 09. I think this is often true. While subject matter of mixed documents may be very similar, the metadata of those same mixed documents could be much dissimilar because of the media the information was created in. Films and video may have many descriptors unique to the media, and differ markedly from wholly print media. Summaries and full-text tend to be quite similar regardless of media.</p> <p># 10. Search systems are not good enough at present. Only be marginal improvements to be had on the system side, but in terms of information in the system, ease of use and ability for the user to customize the search environment and results there is considerable improvement possible.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 16 (Continued)

National Agricultural Library (NAL)	A			X		<p># 09. If “more accurate” can be interpreted as “more comprehensive”, I agree. I don’t think systems designers always take into account all searchable features of items, or that content preparers do the best job of preparing “raw material” for searching.</p> <p># 10. I continue to like using measures of volume and speed of retrieval, precision and recall, as well as usability testing. However, if they can be supplemented with more human-intense follow up, I believe that performance can be tested more fully.</p>
National Archives and Records Administration (NARA)	A	X	X			<p># 09. I think that improving search systems will continue to benefit power users (internal staff, clients, and/or experienced researchers who constitute major stakeholders in a system) and is worth doing if power users have unmet search needs (i.e., they are frustrated with their inability to perform more advanced search functions in the system).</p> <p># 10. User education and virtual support (via email, chats, and webpages providing help and search hints) are means of helping less experienced users to be more successful in their searches.</p>
National Archives and Records Administration (NARA)	B	X	X			<p># 09. Perhaps what are needed are not more accurate search systems, but rather improved search tips and help to guide the searchers in conducting more effective searches.</p> <p># 10. No comment.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 16 (Continued)

National Library of Medicine (NLM) NIH	A	X	X			<p># 09. I don't feel I have enough knowledge to comment on this question. I do think enriched metadata improves results with current search engines.</p> <p># 10. I don't feel I have enough knowledge to comment on this question, either.</p>
Office of Scientific and Technical Information (OSTI) DOE	A				X	<p># 09. Accurate searching implies being inside the searcher's head. Only the searcher knows what he or she wants and sometimes they don't even know which is the discovery part of what we do. Relevancy ranking is where the R&D needs to be focused, as well as the capability to turn it on or off at the searcher's whim....</p> <p># 10. .. People are always trying to compare Google against Yahoo against MSN against Ask etc., adnauseam. There are even folks out there that have built a distributed search that covers all 4 of the aforementioned search engines and have added clustering to boot...</p>
Office of Scientific and Technical Information (OSTI) DOE	B	X				<p># 09. The primitive search systems we have today are a lot better than nothing, but emerging generations of search systems will provide enormous benefits.</p> <p># 10. I am interested in improving search engine performance. I will let other folks do the measuring. For example, you don't have to take quantitative measurements to know that the searching done by Science.gov 3.0 is far superior to that offered by Science.gov 1.0, just 4 years ago. Major further improvements are coming.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 16 (Continued)

USGS Biological Resources Division (Dept. of Interior)	A			X	<p># 09. No, I think often the results are there, but due to rankings, user interface, and user unfamiliarity with the search system, this is why it appears. Results are not accurate or users don't retrieve what they desired.</p> <p># 10. Search engine performance is ultimately measured by 1) has the search query returned what a user expects. This is somewhat flawed in that all of us do not think the same, expect the same results, and/or have different cultural/educational backgrounds.</p> <p>Others try measuring search engine performance simply by:</p> <ul style="list-style-type: none"> • Cost to acquire (Have to consider the life cycle cost of the Engine!) • Collection Size and ability to handle large volumes of data – this is important, but just because an engine can handle over 1 billion documents, does that mean your organization is just adding a bunch of garbage to the engine index. Are those 1 billion documents key documents, can they be parsed successfully, subsisted for users, etc. • Maintenance/Operation Resources Required – over the life of the engine (which is probably no more than 5 years) <p>User Interface customization based on user preferences</p>
---	----------	--	--	----------	---

SEARCH SYSTEMS PERFORMANCE AND MEASUREMENT

Question # 9 & 10

DOD Organizations and DOD Contractors

Table # 17

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p># 9. ...current state of search engines...</p> <p># 10. ...flaws in measuring search systems performance...</p>
Air Force Research Laboratory WPAFB	A		X			<p># 09. They probably are good enough – it's just that there are so many of them. The days of one overarching databank – such as Dialog – serving as an exhaustive federated search tool – are gone. The major sci-tech publishers compete fiercely to develop their own search engines and platforms, and in so doing deny their content to the older transaction-based systems like Dialog. Unless a library contracts for another federated search tool, to try to recreate the “one-search” capabilities of Dialog, scientists are forced to go to several different platforms for an exhaustive search.</p> <p>#10. No comment!</p>
Chemical and Biological Information Analysis Center (CBIAC)	A			X		<p>#09. Again, I should think it would depend on the topic or area to be searched. If I am looking for test results for the effect of VX on polycarbonate materials at low temperatures, for instance, I would benefit from the most accurate system available. If I am trying to survey or identify technologies used in stand-off detection, I would not want to limit those results unnecessarily - I'd want a very</p>

DOD Organizations and DOD Contractors

Table # 17 (Continued)

Chemical and Biological Information Analysis Center (CBIAC)	A (Cont.)			X		<p>inclusive search and would therefore NOT benefit from exquisitely precise searching.</p> <p>#10. I've seen search times which seemed respectable become unacceptably long once the search is expanded to include additional qualifiers - so measuring the search speed should be done under less than ideal conditions.</p>
Chemical and Biological Information Analysis Center (CBIAC)	B	X				<p>#09. Potentially but there is always room for improvement and perhaps solving the concern of locating documents that are assigned low relevancy ranking.</p> <p>#10. No Comment Provided!</p>
Johns Hopkins University, Applied Physics Laboratory	A		X			<p>#09. Librarians know what they are doing! Search interfaces are only marginal! We need advanced search button!</p> <p># 10. Link between users! TREC test data, computer science, needs human intervention. Need real questions with real users!</p>
Johns Hopkins University, Applied Physics Laboratory	B		X			<p># 09. Don't know! Too much recall in full text searching! For example, INSPEC (Electrical Engineer, Computer Science) database indexed in many ways! It helps in post processing; some databases have begun to do so!</p> <p># 10. Who is doing the measuring? How is the data being measured?</p>

DOD Organizations and DOD Contractors

Table # 17 (Continued)

Johns Hopkins University, Applied Physics Laboratory	C			X		<p># 09. No! Improve interface and user interaction. This will improve search results.</p> <p># 10. Would not use recall, instead, judge performance by precision.</p>
Lackland Air Force Base	A				X	<p># 09. Well, anything can stand to be improved. However, when a user or any searcher better understands the system that they are using, the better they can achieve results they expect.</p> <p># 10. When I read some of the reports that compare various search engines, I always ask myself if apples were compared to apples or were apples and oranges being compared. It is similar to the Consumer Reports guides that compare various cars against one another in performance, satisfaction, and service areas. The cars are all engineered to run differently, so is a fair comparison being done? Do the cars all do the same thing? Do the cars all have the same features? Are the features all described using the same terminology?</p> <p>To overcome these issues- you just have to be willing and able to take the time to learn the database/search engine you are using. In the long run it will save you a lot of time and frustration.</p>
MITRE Corporation	A	X				<p>#09. There is clearly always room for improvement in search algorithms but the ideal system is basically impossible because of the fact that a typical search (which is less than 2 words) can often be interpreted by humans in a multitude of ways. How is a search system supposed to be able to figure out exactly what this user is looking for? For example, “IRA” is a common search term entered in the IRS.gov search box. What kind of information does the user want on the topic of IRAs--the yearly limits or how much he can take out per year or how to set one up? What if the acronym “IRA” is rarely used in the content but is instead spelled out fully. One way you</p>

DOD Organizations and DOD Contractors

Table # 17 (Continued)

MITRE Corporation	A (Cont.)	X				<p>address the imperfection of the search results is to tailor the content appropriately to at least provide a good starting point for all of the possible interpretations. Yet again, much of the solution to search problems comes down to understanding information seeking behaviors and providing content to guide the user in the discovery process.</p> <p>#10. As noted above, relevance is very subjective. If several people enter the search query “IRA,” their opinions on the relevance of the top results may vary widely depending on their actual information need. What we have done at IRS.gov is, for general queries of this type, to provide a good landing page as the #1 result with more specific pages lower down on the results page so that, if the user sees a title that fits his information need, he can go directly to it. Note that, if the title is not informative, the user will not recognize that the content is relevant.</p> <p>When we have done extensive testing of several search engines in order to choose one for the site, we tested each with the same set of queries, drawn from the most frequent search terms list as well as known problem queries, and, subjectively, identified an “ideal” set of results. We then calculated very soft precision-recall scores, looking at precision after 1 retrieved document (when a form was requested) and after 5 documents were retrieved (when the query was for general tax information).</p>
MITRE Corporation	B	X	X			<p>#09. Search systems are not good enough! Yes systems need to get better with more relevant results! Both accuracy and usability need to be improved! Google provides searchers with sociability with their search experience!</p> <p>#10. TREC has tried to address this issue! Scalability is an issue! Need to do measures with large data sets. Also, sociability and usability must be measured in any search engines’ evaluation.</p>

DOD Organizations and DOD Contractors

Table # 17 (Continued)

Naval Research Laboratory (NRL)	A			X		<p># 09. Perhaps a scholar should be searching multiple systems regularly. A 2nd opinion is always a better approach. In scholarly research ‘good enough’ is not good enough.</p> <p># 10. I will create a term “metadata stacking”—there might be a better way of expressing this. Strict standards for metadata creation might limit the number of useless or barely useful results that appear in some databases.</p>
Pentagon Library	A	X				<p># 09. There is more room for improvement!</p> <p># 10. The need for more interfaces! The creator and user need to work to together!</p>
US Army Library Picatinny Arsenal, NJ	A	X	X			<p># 09. I think this is true. I can’t foresee any improvement to a full-text/metadata search that would generate better results.</p> <p># 10. With full text searching, you must be able to eliminate unwanted hits; i.e. hits based on citations in the reference, when in fact you just want reports done by a specific author, not his reports that were cited. Therefore you must have a combination of metadata and full text searching.</p>
Redstone Scientific Information Center (RSIC)	A				X	<p>#09. I have no idea about whether this is true.</p> <p>#10. No Comment Provided!</p>

DOD Organizations and DOD Contractors

Table # 17 (Continued)

Redstone Scientific Information Center (RSIC)	B	X				# 09. Agree. #10. No Comment Provided!
---	---	---	--	--	--	---

SEARCH SYSTEMS PERFORMANCE AND MEASUREMENT

Question # 9 & 10

UNIVERSITY PROFESSORS RESPONSES

Table # 18

	PARTICIPANT	F U L L T E X T	M E T A T A R	O T H E R	N O P R E F	COMMENTS
						<p># 9. ...current state of search engines...</p> <p># 10. ...flaws in measuring search systems performance...</p>
Old Dominion University	A	X	X			<p># 09. I believe there is still scope of improving search engines.</p> <p># 10. It is difficult to characterize the user model, that is what</p>
Old Dominion University	B					<p># 09. Well, scholars are not above asserting tautologies. Of course more accurate search systems would lead to more accurate searches.</p> <p>As to whether search systems are good enough, that seems to be highly dependent on the application and the user community involved.</p> <p># 10. The traditional measures of precision and recall are based fundamentally upon a notion that documents' relevance is <input type="checkbox"/> Boolean rather than ranging over a wide variety of possible relevance strengths. A better measure would require a proper statistical model of the uses made of retrieved documents.</p>

UNIVERSITY PROFESSORS RESPONSES

Table # 18 (Continued)

Syracuse University	A		X		<p># 09. ... I wouldn't expect major gains to be made, in say, expert medical searching or legal searching where vocabularies are very rigid and the users conversant in the content matter. Searching for music, however, has taken leaps forward recently, mostly by bringing old fashioned metadata techniques to the field. There are HUGE strides needed in both the multimedia and spatial worlds.</p> <p># 10. Not clearly defining the metrics being used and the outcome being measured. Know thy users.</p>
Syracuse University	B	X	X		<p># 09. I am a believer of metadata and ontology supported information searching systems. Full-text search can only do this much and often needs to be used together with metadata. The technology is sophisticated enough now to provide good search results, but scholars still feel the systems are not good enough. The real reason is the absence of semantic infrastructure now to provide good search results, but scholars still feel the systems are not good enough. The real reason is the absence of semantic infrastructure – mapping between controlled vocabulary and keywords that will point users from one to the other no matter where they start a search.</p> <p># 10. Not familiar with this topic.</p>

UNIVERSITY PROFESSORS RESPONSES

Table # 18 (Continued)

San Jose University	A		X			<p>#09. No.</p> <p>#10. N/A</p>
University of North Carolina	A			X		<p>#09. Not sure what ‘accurate’ means----today’s Google is much better than the Google of 3 years ago---some of this is corpus-based (better crawlers, more link structure, more documents, etc.), some is engineering based (better caches, networking), and some is search algorithm (human tuning of the SE takes place on a daily basis)</p> <p>#10. The primary flaw: Assuming that one measure fits all IR contexts or tasks.</p>

SEARCH SYSTEMS PERFORMANCE AND MEASUREMENT

Question # 9 & 10

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 19

	PARTICIPANT	F U L L T E X T	M E T A D E T A	O T H E R	N O I S E	COMMENTS
						<p># 9. ...current state of search engines...</p> <p># 10. ...flaws in measuring search systems performance...</p>
Access Innovation Inc.	A				X	<p># 09. The search software itself is pretty good. The presentation of the results and the options to access the corpus need a lot of work. We only provide one way in to the data.</p> <p># 10. Relevance, precision, and recall are each measured subjectively by a human. We assume there is only one valid answer set. We allow only one way to search usually – either a single or a series of boxes. Search results also vary by the user expectation. What is actually in the file?</p> <p>I think another way to measure the results is HITS (those a human thinks are appropriate) MISSES (those a human would chose and the system did not) and NOISE (those the system chose and the system did not) NOSIE can be both relevant and irrelevant depending on the level of expertise of the human reviewing the material.</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 19 (Continued)

Information International Associates Inc.	A	X	X		<p># 09. This totally depends on the domain and context. Sometimes the results are too much so is that too good or is that not good enough? A cost benefit trade-off of more accurate systems vs. more than marginal improvements depends on the specific context.</p> <p># 10. Precision and recall as far as I know still require expert opinion, so there are some flaws in that process.</p>
Information International Associates Inc.	B			X	<p># 09. . . . it depends on what the searcher wants. The question of what is “good enough” depends on the reason for the search. Certainly one could argue that when dealing with life or death situations “good enough” doesn’t cut it. If you are looking for a place to get started and want just a few documents, then you don’t need as “accurate” a search engine.</p> <p># 10. There have always been flaws in the process. Search engine performances (if you are talking from the results side only) are geared toward the traditional recall and precision. These have always been difficult because they are dependent on the user, the question, the context, etc. I think it probably also depends (as does indexing) on the stage of the moon.... What is good to a user one day may not be good to the user on another day.</p> <p>I think one way to overcome these issues is to provide good help and suggestions so that people can try different “methods” of searching for the same item. It is often helpful too to ensure that both search and browse approaches are available. A third approach is to provide different paths into the same document base (this is often done through metadata or faceted controlled vocabularies that are reflected in the taxonomy). It can also be done by providing links within documents that execute searches</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 19 (Continued)

Information International Associates Inc.	B (Cont.)			X		back to the database. For example, in the ERIC database, you can do a search, click on a result, and the author name will be highlighted if there is another document in the database by that author. When you click on the highlighted author it executes a search on the author name and you have immediately broadened beyond the confines of the initial search and its limitations.
National Federation of Abstracting & Information Services (NFAIS)	A				X	<p># 09. In an ideal world, users would search for content in environments that supported various learning styles, various community practices and a full range of formats. We may never reach that ideal environment. Search support is "good enough" when a critical mass of users is satisfied with the quality, depth, and the amount of the information that they retrieve. In some situations, we're there now. In other contexts, we're not anywhere near the benchmark of adequate performance.</p> <p>#10. The biggest problem is ambiguity of language which can be countered to some extent by controlled vocabulary and other mechanisms for refinement of queries. However, another significant problem that is not currently being addressed is making known the scope of the content available for searching. It would seem to me that soon (within the next 18-24 months); users will begin to recognize this as an issue. They will be working from expectations formed in a world of Flickr, iTunes, Blogger, YouTube, etc. and other environments specific to their workflow (such as the Virtual Observatory in the field of astronomy). They will go exploring for full text books in Google Books and wonder whether the information environment provided in the workplace has the same functionality. Users will be saying to themselves "I can find all of this for free; can I work this way at the office?" and because it will be work-related they'll be concerned</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 19 (Continued)

National Federation of Abstracting & Information Services (NFAIS)	A (Cont.)					with whether everything will be included. Even worse, they will <i>assume</i> that they have access to everything they require and blame information providers when they find the gaps in coverage. Note that this will apply to subject areas, traditional and non-traditional formats.
National Commission of Libraries and Information Science (NCLIS)	A	X				<p># 09. We can always improve a system but we may not see just how to do that today.</p> <p># 10. I am not sure but knowledge of what the search engines do would be a start.</p>
Southeastern Library Network (SOLINET)	A				X	<p>#09. I believe all electronic search systems are incomplete. So, live long the books in the stacks!</p> <p>#10. That search engines are accurate and all the information can be found on the web.</p>

SEARCH SYSTEMS PERFORMANCE AND MEASUREMENT

Question # 09 & 10

OTHER LIBRARIES RESPONSES

Table # 20

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p># 09.... current state of search engines...</p> <p># 10. ... flaws in measuring search systems performance...</p>
Catholic University of America	A			X		<p># 09. I don't know. Algorithms that deal with common misspellings are useful.</p> <p># 10. I don't know.</p>
Senate Library	A		X			<p>#9. I'm sure improvements are quite possible. However, sophisticated systems that will automatically assign lots and lots of metadata tags to incoming content (this would improve results) are expensive and take a lot of expertise to set up and maintain. There has to be an institutional commitment for this. I also think that bibliographic instruction (or whatever it is called these days) is vital, or else people just flounder around, or think that what they find on Google or by a cursory search of ProQuest is "good enough" or even worse, that the cursory search is exhaustive! I think that incoming students at universities should be required to take at least one course in bib instruction. That goes for faculty as well.</p>

OTHER LIBRARIES RESPONSES

Table # 20 (Continued)

Senate Library	A (Cont.)		X		#10. I think that a big problem is determining whether people find what they *really* wanted ... or even more, that they found something that they weren't originally looking for, but that actually gave them better information than they had realized existed. (Sorry about the convoluted syntax, but I hope my meaning is clearer than my prose!) See my comments above on cursory searches. I don't know how one overcomes those issues—those issues existed in the days of the card catalog and the printed index.
-------------------	--------------	--	---	--	--

SYSTEM IMPROVEMENTS AND RETRIEVAL EFFECTIVENESS

Question # 13, 14, 15, & 16

CENDI MEMBER AGENCIES RESPONSES

Table # 21

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 13, 14, 15& 16. System improvements, data retrieval effectiveness and barriers to the user search experience.
Defense Technical Information Center, (DTIC)	A				X	<p>Lots of research on different alternatives</p> <p>Process and handling large amounts of data to get decent response time! Need better taxonomy to improve accuracy. Multiple thesauruses! Ability to drill down to get better results!</p> <p>Multiple thesauruses and drill down capability.</p> <p>Content sensitive! Make things simple! Speed would be a benefit to increase response time. Processor speed is holding us back. Home bandwidth has limitations to the user. This places limitation on downloading capability.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

Defense Technical Information Center, (DTIC)	B			X	<p>Try to provide alternative searching capabilities for the user to have available!</p> <p>Web search engines handle a lot of data but don't have much precision. They try to improve that by improving the relevancy of the articles on the top of the hit list. They are good for one good article on something. But they don't search the deep web or for-pay databases. More metadata or the ability to search within a sentence or paragraph would help, rather than on the whole document. But these systems purpose is mostly to answer easy questions, find resources, shopping, etc., which they are good at. They aren't set to find all data on a subject, as researchers and patent attorneys want. So basically they are as effective as they need to be for their purpose, until there is so much information on the web that they can't handle it.</p> <p>Need for time, patience and knowledge: I don't ever expect busy professionals to do their own sophisticated searching, where they need to know what sources to use, the content of the collections, and the peculiarities of each collection/search interface. For that, you still need intermediaries or other expert searchers to provide added value.</p> <p>Intermediary! Time and patience! Familiarity with subject and collection!</p>
---	----------	--	--	----------	---

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

Defense Technical Information Center, (DTIC)	C	X				<p>There will be large-scale improvements. In the near term as more documents are created using a common XML meta tag structure and there is an metadata of imbedded into documents as they are created search crawlers will better understand the data they are indexing, improved searching will result. I expect that with ever increasing CPU cycles per server, search engines will be able to derive content and content from unstructured data.</p> <p>The biggest limitation is relying on searching. You may miss the document that you are looking for because of flaws in the database. Documents may not be put in the database correctly which leads to poor search results. There is a need for other mechanisms for cataloging to ensure that you have retrieved all your documents. May need to do document by document review.</p> <p>By segmenting the collection to improve your search results. Subject categories. There need to apply a broad thesaurus across specific categories of content will gives searchers additional clues and options.</p> <p>There are not enough human factors in building interface. Developed for good searchers, but not designed for the novice searcher. Language will become an issue as the percentage of Americans speaking English as their primary language declines.</p>
Defense Technical Information Center, (DTIC)	D	X				<p>Organizations that still search the bibliographic record can easily make large scale improvements to get up to the level of a Google. But for the Goggles of the world, probably “The low hanging fruit has been picked off”. Personalized searching might make a big improvement, i.e. the search engine is somehow intimately familiar with the types of things that you are looking for, i.e. are relevant to you.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

Defense Technical Information Center, (DTIC)	D (Cont.)	X			<p>Is accuracy really an issue with full-text searching? If the term being searched is found in the document, the search engine will find the document. Relevance ranking, categorization, the search interface, etc are the true keys.</p> <p>Powerful relevance ranking algorithms, better search interfaces, better displays of ‘hits’ such as categorization tools, visualization tools, etc.</p> <p>Users have great expectations that whenever they do a search, that the result will be more ‘Google like.’ Google search results have become the standard by which user expectations are based. Improvement in user interface will also improve the user experience. The use of categorization tools is a boost to search results.</p> <p>Some metadata search systems do not display their controlled vocabulary where it is obvious to the user to improve their search results, by being interactive; instead it is left to the user to determine that there is such a tool. This can be counterproductive when one considers the time, cost and effort in maintaining a controlled vocabulary.</p>
Defense Technical Information Center, (DTIC)	E		X		<p>Yes, however, the information in a search system has to be input correctly. Cataloging and indexing effectiveness play a critical role in the success of any search system’s success. It would be nice to have the capability to search all formats equally. For example, without good metadata on PDF files, they can not be searched effectively.</p> <p>For one, we need improvement in OCR (Optical Character Recognition) results. Because of the time element, OCR software is used to translate images to searchable text, but some words are incorrectly changed in the OCR process. More emphasis needs to be placed on quality control.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

Defense Technical Information Center, (DTIC)	E (Cont.)				<p>We always need more training for users of the various systems!</p> <p>The user's inexperience, lack of knowledge, and lack of training! Users could benefit if search systems had the help information up front and readily available to aid users with each section. Suggestions on the hit list or the ability to refine searches would also be good.</p>
Defense Technical Information Center, (DTIC)	F	X	X		<p>Yes! I expect improvements! It is great that access has improved! Search engines providers will need to improve their systems in order to maintain user interest and to stay in business!</p> <p>How information is presented to the user is important. Also, good user interface.</p> <p>By providing search instructions! Having good explanation that people can understand!</p> <p>Most users have time constraints! If the search experience is difficult, then users will move on! Good tools are important! Users will get frustrated if results are hard to find! Giving the user the ability to place their idea into a search experience! Users obtaining support from intermediaries, such as the library, for assistance!</p>
Defense Technical Information Center, (DTIC)	G			X	<p>No. Unless they invent a computer that understands language as well as humans, we will always have the same tools we have now. Each of these tools can be improved but none will lead to a quantum leap in retrieval effectiveness. The reasons for this are 1. all languages are continually changing, 2. each person describes things differently, both what they do and what they seek, 3. people are creative. Some may think there is a magic algorithm that will always find the best items on a topic.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

<p>Defense Technical Information Center, (DTIC)</p>	<p>G (Cont.)</p>			<p>X</p>	<p>However, algorithms are just equations that second guess the searcher's intent and assume that one searcher is like any other. Others may think that automatic term expansion improves results. A review of how this looks in some of our current products shows how inadequate the systems' understanding of language is.</p> <p>I believe our Boolean search systems are accurate. The systems themselves are fine, but the data quality, interfaces, and controlled vocabulary could be improved. The algorithmic search engines used on the web also seem to be "accurate" enough for their purpose, which is only to find approximate results.</p> <p>Improve the interface design and data quality. Full text databases have been improved by adding fields and controlled vocabulary. But I don't believe the reverse is often true. Adding full-text does not increase relevancy of results. If you have a bibliographic database and then offer the option to search the full-text, most novice users opt for the full-text and are inundated with irrelevant hits where their terms are mentioned only in passing. When full-text is offered a) it shouldn't be promoted as a replacement of field searching, b) it should have some algorithmic ranking, and c) it should also be available for straight Boolean searching.</p> <p>Are we differentiating between recall, precision, accuracy, content retrieval, retrieval effectiveness, and improving user search results?</p>
--	-----------------------------	--	--	-----------------	---

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

Government Printing Office (GPO)	A			X		<p>I believe that there is a progressive growth taking place on both the technological and user sides, but I am unclear as to the magnitude of the improvement this will bring forth. A user's ability to correctly interpret the results they get from a search is every bit as important in determining accuracy as the search itself.</p> <p>Interoperable categorization that can easily be understood and used by the common searcher and readily accessible training tools for them to teach themselves in using the available tools.</p> <p>A lack of understanding by users regarding how information is published electronically and rampant inconsistency in the construction of data and indexes are both severe barriers to useful searching.</p>
Government Printing Office (GPO)	B	X	X			<p>I think making better use of controlled vocabularies will assist here.</p> <p>Keyword and Boolean query based systems; limitations in natural language analysis; limitations in dealing with unstructured languages.</p> <p>Recognition of objects regardless of spatial orientation. Combining improved metadata searching (mentioned above) with natural language searching so they don't operate in isolation but are synchronous.</p>
Library of Congress	A	X				<p>Don't know! Google's real contribution is making people think that they can find what they are looking for.</p> <p>Could add value to formal system. I believe individuals can do their own searching.</p> <p>Convince people that they can get information. Get what they are looking for!</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

Library of Congress	A (Cont.)	X				<p>If the system become more interactive, then searchers will get better results</p> <p>If searchers can clarify what they are looking for, think support other case?</p> <p>When people talk with others about what they are doing, they almost always improve their search results.</p> <p>Lack of clarity in the mind of the searcher.</p>
Library of Congress	B	X	X			<p>There are some visual search engines that use clustering/ mapping.</p> <p>One search engine fits all usually does not fit all- I'd like to see more customization</p> <p>Take time to learn how system works, ask for help from an expert, try multiple search engines, and use controlled vocabulary...</p> <p>I have witnessed researchers who are afraid of searching the Web or database...maybe if there were some kind of personalization or customization researchers would be more at ease...wonder if there were standards for search systems terminology! For example, some database use journal title, source, etc...</p>
Library of Congress	C	X	X			<p>No! Not as long as human beings are the ones doing the searching.</p> <p>Better understanding of users' needs and abilities and ways of searching. Cognitive psychology.</p> <p>Language, fear, general state of mind, mental illness, physical distractions,</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

Library of Congress	C (Cont.)	X	X			<p>attitude of user, design of search system, physical disabilities, level of education, etc. etc.</p> <p>All of these barriers can be effectively minimized by a compassionate and skilled intermediary.</p>
NASA Scientific and Technical Information Program	A			X		<p>The systems themselves will become more efficient, the information in those systems will be much greater and in more variety, and the end users themselves will become more proficient in using the systems.</p> <p>...more machine intelligence built into the search tools, and the ability of the systems to learn from previous use of those systems by users.</p> <p>...systems themselves need to be more focused on the users, and more easily customizable by the users.</p>
National Agricultural Library (NAL)	A			X		<p>Yes, I am confident that we'll figure out how to combine large data sets such as GIS and genomics data with full text and bibliographic searching, for rapid, simple-looking searching.</p> <p>Information professionals often care more about accuracy of searching, while some or many searchers may care more about quantity. Therefore, I would think that effectiveness is "in the eye of the searcher" and if we get good marks from our customers for our systems, which are the most important gauge.</p> <p>Language; bad presentation; poor technology. Ways to overcome usability testing.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

National Agricultural Library (NAL)	A (Cont.)					I think the “suggested terms” for users probably does improve user satisfaction with their search results. As I have mentioned in other responses, I think user education, both formal and informal, is a key way to improve user satisfaction and users’ search capabilities.
National Archives and Records Administration (NARA)	A	X	X			Withheld or buried information – the absence of explanations of how the search system works (i.e., fielded or full-text searching, or a combination), examples of search strategies, explanations of how search results are ranked, and a glossary of specialized terms – is the greatest barrier to a user’s search experience. This type of information needs to be presented in a clear, effective manner that speaks to different types of users at different levels. New users need guidance. I also feel that minimally (or poorly) populated metadata and inconsistently (or inaccurately) applied controlled vocabulary indexing can be barriers to successful search experiences. The quality of the data affects the quality of the search experience.
National Archives and Records Administration (NARA)	B	X	X			<p>There seems to be competing interests in better speed versus better accuracy. Improved accuracy and complexity needs to be achieved without sacrificing search speed or performance.</p> <p>Better search tips/help files Better metadata describing the content of the records</p> <p>One barrier is poor design of the website or database. This can be minimized by conducting usability studies and following industry best practices in user interface design.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

National Library of Medicine (NLM) NIH	A	X	X		<p>I do not expect large scale improvements in the current means of searching. As the web grows, it becomes harder and harder to find selected pieces of text on a page. If search engines begin taking controlled and quality metadata into account, the creators of resources will begin to provide quality, controlled metadata. As it stands, there isn't enough quality metadata available for most search engines to pay attention to it, and not enough search engines look at it for data providers to invest the time it takes to create it. Search engines also need the ability to search controlled vocabularies and map unauthorized terms to those authorized by the vocabularies at the same time to improve search results.</p> <p>Other means of searching, including OAI harvesting and the semantic web, will have far superior retrieval than current online searching, but I do not expect these to take off in a big way. Some institutions will jump into this and stay with it, including libraries with large online image collections, but most organizations may never learn or care about these methods of describing and exposing data.</p> <p>Users also are notoriously impatient and unwilling to scroll; information that is not visible on the first screen is called "below the fold," and there is a very strong psychological resistance among users to look at this information. Unfortunately, there is little the web itself can do for the user that is not willing to help him or herself. Tutorials are very helpful</p> <p>Another problem is user behavior, but there is little that can be done about that. Users avoid clicking on parts of a page that appear to be ads, whether there is advertisement on a page or not.</p>
--	---	---	---	--	---

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

National Library of Medicine (NLM) NIH	A (Cont.)	X	X			<p>Lack of user testing often leads to problems with the finished product. Often what seems logical and obvious to designers is not clear to users. Testing and redesign is key to creating good user interfaces.</p>
Office of Scientific and Technical Information (OSTI) DOE	A				X	<p>...need to employ more parallel processing architectures and use distributed processing. This is the next huge step forward in the information business.</p> <p>...we are now approaching some processing barriers if we want to employ robust relevancy ranking and still retain a fast response time. Once we take the step forward with more distributed processing and use parallel processors to do the relevancy ranking utilizing more powerful and encompassing algorithms we will see major advances in science.</p> <p>Yes, more powerful relevancy ranking tools. Clustering is nice, as are images, but the heart of the matter is with millions of new information resources being created each day you need toolsets that can take advantage of these new resources and bring the best and most accurate information to the searcher in the shortest amount of time possible.</p> <p>The biggest ‘barrier’ for me is the amount of information available I generally look at the first 30 hits size them up and pick the ones that seem close first.</p> <p>...discovery is a huge part of the process when you search, learning to better scope what you do or don’t want. Finding a trail you did not know existed before and following it. That is why the user needs the ability to turn the relevancy ranking tool off if they want.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

Office of Scientific and Technical Information (OSTI) DOE	B	X				<p>We at DOE intend to develop new and improved search systems and to be the first adopters of advances that other developers make.</p> <p>Metasearch has enormous potential that has yet to be realized. Relevance ranking in a distributed environment is still in its infancy, and it will mature rapidly over the next several years.</p> <p>Better engines, better relevance ranking algorithms, and improved precision search tools in general.</p>
USGS Biological Resources Division (Dept. of Interior)	A			X		<p>Yes, with Natural language improvements, pattern recognition, inference, and semantic technologies improvements should occur. However, most of these improvements are still based on having some sort of metadata or high quality information about the document and item. Improvements need to be made in the creation of this metadata and in quality control for these efforts to fully succeed</p> <p>The lack of comprehensive vocabularies, fully understanding diverse user requirements, simple yet powerful user interfaces, and the overall volume of non-relevant data/information are huge issues. Research is ongoing in several of these areas and will help to address some of the issues; however, until search tools can read users minds as to what they really meant, 100% satisfaction will not be achieved.</p> <p>Define your user groups, usability studies, simple user interface, more powerful (often behind the scenes vocabularies), recognition that improving search results is a full-time multi-disciplinary position (IT, KM, Domain Expert) that requires dedicated resources, recognition that technologies/tools are always changing and this is not a reason to jump on</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 21 (Continued)

USGS Biological Resources Division (Dept. of Interior)	A (Cont.)			X	<p>the latest tool on the market, understanding the content that is being served by the search engine, understanding the tools limitations and strengths, eliminating government “ease” when building such systems, and finally putting more power (not necessarily choices) in the hands of the users.</p> <p>Cultural, the tool itself, unsure of the content that it is supposed to retrieve, lack of understanding of the domain of the content, time available to the user to fully read/digest Tips/Help/Scope of the Index, previous impression or where they successful or not, response times, and the volume of seemingly unrelated information.</p> <p>This can be minimized by focus groups, flexible user interfaces, and acknowledgement that search tools are a vital part of the business and should have the necessary resources.</p>
---	----------------------	--	--	---	---

SYSTEM IMPROVEMENTS AND RETRIEVAL EFFECTIVENESS

Question # 13, 14, 15, & 16

DOD Organizations and DOD Contractors

Table # 22

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 13, 14, 15& 16. System improvements, data retrieval effectiveness and barriers to the user search experience.
Air Force Research Laboratory WPAFB	A		X			<p>With clustering technology – they may not have to. See clusty.com</p> <p>Better training – librarian-developed and provided.</p> <p>Too much available – users are confused. Federated search tools such as CSA's Multisearch are something we have tried.</p>
Chemical and Biological Information Analysis Center (CBIAC)	A			X		<p>Don't know.</p> <p>Ability to combine and manipulate search sets.</p> <p>Ability to review the search results in a bit more detail - on screen output could be designed differently (?dynamically) from the output formats used to generate bib files.</p>

DOD Organizations and DOD Contractors

Table # 22 (Continued)

Chemical and Biological Information Analysis Center (CBIAC)	A (Cont.)			X		<p>I would like to be able to do a broad search, select a subset of the broad search and then print out the selected records and the non-selected records in 2 different bibs. One would have the freedom to do a very precise search but then to present a secondary bib. This would require being able to produce user-defined sets and to manipulate those sets.</p> <p>It's frustrating when searches time out.</p> <p>It's frustrating when a complicated search strategy fails but then cannot be recovered for review and tweaking.</p>
Chemical and Biological Information Analysis Center (CBIAC)	B	X				<p>Allow searching using limited distribution; adjust searching in an advanced mode to allow for extended searching in the various volumes.</p> <p>Helpline with a human versus a computer/recording.</p>
Johns Hopkins University, Applied Physics Laboratory	A			X		<p>Facet search results...coming from results sets...takes author name associated and group it! Also clustering!</p>
Johns Hopkins University, Applied Physics Laboratory	B			X		<p>Don't expect much improvement! I don't see any large scale improvements! From a searching capability, users can improve their skills! Post processing!</p> <p>Information storage problem! Information that could be searched (to provide value) requires so much storage! Data may not be available. Bandwidth necessary to access information may not be available to all users!</p>

DOD Organizations and DOD Contractors

Table # 22 (Continued)

Johns Hopkins University, Applied Physics Laboratory	C			X	<p>No! No return on investment! Not hard science! Hard to sell! Funds not allocated due to lack of support from top management.</p> <p>IT personnel don't understand how the average person searches a database. Need to improve the amount of interface.</p> <p>Don't let IT folks design interfaces. Really pay attention to what the users need and what they are generally looking for.</p> <p>Utilize as much "controlled vocabulary" as the budget will allow, i.e., cost of developing the controlled vocabulary and the cost of indexing.</p> <p>Provide a staff member whom users can contact for some human interaction. User questions are the best feedback.</p> <p>IT developed interfaces are the number one barrier. They are not intuitive to the average, or even sophisticated searcher. IT folks think in a different way than the rest of us. And no matter how good the search structure, taxonomy, etc., if the interface is bad, the user will never find out the other good stuff.</p>
---	----------	--	--	----------	--

DOD Organizations and DOD Contractors

Table # 22 (Continued)

Lackland Air Force Base	A				X	<p>No. Google has taken over and dominates the search engine scene according to users and Googles stick rates. Unless the leader of the industry determines that users are not getting what they need, not much will be done to improve accuracy and retrieval rates. Users do not consider this an issue. Users do not know that they do not know enough about researching. Users think that anyone can research. I face this in my library daily.</p> <p>The users themselves who feel that anything can be found on Google. The methods used to compare various search engines. The way that search engines describe their capabilities. If some standardization were possible to be achieved in the industry, everyone could be reading from the same page of music.</p> <p>Training. Education.</p> <p>Users do not know how to select useful and relevant search terms. Users do not understand how to search. Users do not learn how to use various databases.</p>
MITRE Corporation	A	X				<p>I don't really know.</p> <p>As noted in previous questions, better understanding of the role that information seeking behaviors play in the success of search experiences is critical. A holistic view that does not just focus on the search engine's results but instead looks at the whole user experience will make the difference.</p> <p>Understand the user population (How many are novices or new to the website, which will determine how familiar they might be with the navigation and the terminology used on the website? What are they</p>

DOD Organizations and DOD Contractors

Table # 22 (Continued)

MITRE Corporation	A (Cont.)	X				<p>seeking when they come to the website? Do the majority need high recall or high precision?). At IRS.gov, we have worked to improve the whole user experience by making sure that common terms that people enter as search terms that may be found in the content retrieve reasonable results; we added informative titles in the search results displays for documents whose title metadata tag did not supply good titles; we highlight “recommended results” (quick links); we use the search thesaurus extensively to force prime results to the top of the search list. In addition, as noted in the next question, we constantly review the frequently entered search terms list to capture variations on common form names. Why should the user be required to know the exact form number title if we understand what he is looking for?</p> <p>Search terms may not match jargon or business-specific terminology. Content may include a number of synonymous terms, depending on the author, where uniform use of terms would be better.</p> <p>Acronyms used in the content may not be familiar to users.</p> <p>Users may use a variety of ways to search for the same content, for example, on IRS.gov, we have identified more than a dozen search terms equivalent to the 1040-EZ form (e.g., ez1040, 1040ez, 1040 ez, form1040ez, e-z).</p>
-------------------	--------------	---	--	--	--	--

DOD Organizations and DOD Contractors

Table # 22 (Continued)

MITRE Corporation	B	X	X		<p>Expect improvement in audio and video, they are both poor today! The need is there! Also, language processing, real shift 10-15 years, as performance systems improve. TREC language processing to improve retrieval has not resulted in improvement!</p> <p>There will always be the issue of human interaction that breeds inconsistency, for example in indexing! With machine aid indexing will reduce cost and time, but it will not be as good a human beings.</p> <p>Usability issues! Systems do not interact well, documents versus multi-media! Inaccuracies in search systems! Cataloguing is a problem!</p> <p>We have not yet figured out the most effective search interface! There is the need to help the user formulate searches for better results. A need for commonality across search systems to allow for the exposure on information space to improve search results! Need usability of system with post retrieval exposure!</p> <p>User's unwillingness to specify searching needs. Poor user selection of search terms! Lack of experience in searching! Through training and education a searcher experience will improve!</p>
--------------------------	----------	----------	----------	--	--

DOD Organizations and DOD Contractors

Table # 22 (Continued)

Naval Research Laboratory (NRL)	A			X	<p>Unless a completely different approach is taken, I do not foresee any large scale improvements. I need a crystal ball for this one.</p> <p>The modern element of information retrieval impatience is one problem that needs to be addressed. If the good results are not produced first and quickly, younger researchers may not choose to take the time to search deeper and longer. The deep net/hidden net needs to be more fully explored and better ways developed to utilize information hidden there.</p> <p>Teaching better strategies! “For Dummies” help tips that are tested by young searchers.</p> <p>In some environments, there are users who are techno-deprived. In some cases, very basic needs must be addressed.</p>
Pentagon Library	A	X			<p>Hope for improvement in the accuracy retrieval rate. An increase in information!</p> <p>A means to create a search system with taxonomy! Users not thinking in terms of the way subject headings were created. More natural language! More acceptance in satisfying the common user!</p> <p>By using full text! It is easier!</p> <p>By making searching easier! If accurate metadata is assigned, then search results should improve!</p>

DOD Organizations and DOD Contractors

Table # 22 (Continued)

US Army Library Picatinny Arsenal, NJ	A	X	X			<p>No, not until the 'junk' is removed form the internet. I know that 'junk' will be defined differently for each person, but I think serious scholarly researchers would like the twin internet system.</p> <p>Get rid of the 'junk'. Insist on metadata for all documents on the web.</p> <p>Get rid of the 'junk'.</p> <p>Lack of knowledge to controlled vocabulary is a hindrance. There should be an online thesaurus for any database with controlled vocabulary.</p>
Redstone Scientific Information Center (RSIC)	A				X	<p>Give them more fields to add terms and narrow the search to make it more focused.</p>
Redstone Scientific Information Center (RSIC)	B	X				<p>No Response!</p> <p>I would like to be able to do a broad search, select a subset of the broad search and then print out the selected records and the non-selected records in 2 different bibs. One would have the freedom to do a very precise search but then to present a secondary bib. This would require being able to produce user-defined sets and to manipulate those sets.</p> <p>It's frustrating when searches time out.</p> <p>It's frustrating when a complicated search strategy fails but then cannot be recovered for review and tweaking.</p>

SYSTEM IMPROVEMENTS AND RETRIEVAL EFFECTIVENESS

Question # 13, 14, 15, & 16

UNIVERSITY PROFESSORS RESPONSES

Table # 23

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						# 13, 14, 15& 16. Retrieval Effectiveness and Barriers
Old Dominion University	A	X	X			<p>Yes, with the use of large computing power available and innovation in parallel algorithms.</p> <p>Lower precision</p> <p>Incorporate semantic searching.</p> <p>Improve precision and classify result sets.</p>
Old Dominion University	B				X	<p>For internet-wide searching, probably not, as there appears to be no prospect for moving people away from WYSIWYG visual formatting of documents to logical/structural markup</p> <p>Incorporate more secondary sources or information.</p>
Syracuse University	A		X			<p>Yes. With the advent of both semantic computing, and the increase in simply the brute force that can be brought to searches they should</p>

UNIVERSITY PROFESSORS RESPONSES

Table # 23 (Continued)

Syracuse University	A (Cont.)		X		<p>improve. Many of today's search algorithms were developed in an environment of computing scarcity. Now we can play with more inductive and heuristic systems.</p> <p>The ability to cross domains in searching and better synthesize results. Over that 50 years the idea was to get a lot of good documents on a topic, now it will be not to simply get the "best" documents, but some picture of how ALL the documents fit together.</p> <p>Listen to the user.</p> <p>Interfaces can get better, and more integrated into user workflows, not simply as a stand alone system waiting for a user to stop what they are doing and go to the search engine.</p>
Syracuse University	B	X	X		<p>The trend for research and education information is to make them searchable through internet search engines. This is a great benefit for scholars and learners. Although the recall has a lot to be desired, the precision has been pretty good in Google scholar. The combination of internet search engines and commercial databases and library OPACs will probably improve information retrieval on a larger scale than ever before.</p> <p>The classic dilemma is how both recall and precision can reach the high level at the same time. After 50 years of information retrieval research, this problem seems to have remained unchanged.</p> <p>Developing ontology's for domains and mapping keywords and controlled vocabularies.</p>

UNIVERSITY PROFESSORS RESPONSES

Table # 23 (Continued)

Syracuse University	B (Cont.)	X	X			Too many to name them. There are different user groups and their information searching literacy levels vary greatly. There is no way to talk in the general term since they will be biased and incomplete.
San Jose University	A		X			<p>No. More different formats of information will make the task more difficult.</p> <p>In the past 50 years, information literacy wasn't paid too much attention. It becomes more important now.</p> <p>Educate the users. Information literacy should be incorporated to school/university curriculum.</p> <p>When people become information literate, the barriers will be minimized.</p>
University of North Carolina	A			X		<p>Of course...passages; multimedia; cross-language.</p> <p>Get people to use relevance feedback</p> <p>Query articulation---create UIs that encourage longer, more detailed queries</p>

SYSTEM IMPROVEMENTS AND RETRIEVAL EFFECTIVENESS

Question # 13, 14, 15, & 16

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 24

	PARTICIPANT	F U L L T E X T	M E T A T A	O T H E R	N O P R E F	COMMENTS
						# 13, 14, 15& 16. Retrieval Effectiveness and Barriers
Access Innovation Inc.	A				X	<p>Better application! Adding controlled terms and allowing use of all synonyms in search. Providing several ways to search so that most learning styles and cognitive processes are accommodated. Using the controlled vocabulary to expand the search query as well as to apply metadata to the records – use it at both ends – see MAIQuery or use it at www.mediasleuth.com. Keeping the controlled vocabulary abreast with the changes in the field.</p> <p>Size, depth and breadth of coverage are not the problems today. Problems have to do with different uses of the same terms by different groups. Accommodating the vernacular is the big problem/ that is disambiguation of terms effectively and of course allowing different ways of access to the data.</p> <p>Presentation of results, manner in which search is allowed. These are not really hard changes to make. We have the tools at hand. We just haven't executed them.</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 24 (Continued)

Information International Associates Inc.	A	X	X		<p>Yes, there is a lot of brainpower and resources, both government and private being invested in Search today. It's a very visible area of research and development. I don't have expertise in where the break through will happen or even if it will be more brute force investment in incremental changes that will make the difference, but I believe consumer demand will drive it.</p> <p>Vocabulary control and subject switching – not new. Better semantic understanding.</p> <p>Clearly user education although that is very hard. One can look to NLM to see some good paths but they are expensive. UMLS where other words and concepts are suggested and this improved user results. I also believe that clustering and visualization is the future to help people hone searches.</p> <p>Too many search results is always a problem and that has not been solved in today's environment. Relevance ranking is one approach that is being used and can be improved. Cluster and visualization can also help here.</p> <p>Understanding the quirks and details of each search system can be overcome by training and good help options.</p>
---	---	---	---	--	---

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 24 (Continued)

<p>Information International Associates Inc.</p>	<p>B</p>			<p>X</p>	<p>I do think that ontology's when implemented through semantic web tools will help to make certain types of information more retrievable. I think that the most benefit will be seen in small, specialized domains. Also, retrieval effectiveness will be helped by the further development of portals, customized environments and retrieval systems that "learn" from the users experience what he or she wants. Of course, the problem is that they aren't always looking for the same thing.</p> <p>I think the traditional Boolean search needs to be combined with more effective semantic/concept-based searching. By that I mean the ability to search for not only terms but how they relate to one another. This requires ontological structures sitting behind the search engine. We need good tools to turn our current tools, like thesauri, into richer structures. We need subject matter experts to help in these areas as well, since they can also help by building these structures in the front-end.</p> <p>I think one of the biggest barriers to a user search experience is the lack of time a user is willing to spend on a search. There is a big difference between the end user and a trained searcher. It has to do both with where their time should be focused and also the fact that they don't know what they don't know. This problem has been experienced time and time again by those trying to teach the use of electronic resources to undergraduate students. The instant gratification is a problem. One of the benefits of growing up in a more paper-based research process is that you learn patience.</p>
---	-----------------	--	--	-----------------	--

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 24 (Continued)

National Federation of Abstracting & Information Services (NFAIS)	A				<p>X</p> <p>Industry experts differ on whether processor speeds will continue to improve so technology may not be the path to improved retrieval. Incremental improvements can be made based on how users behave in their information seeking tasks. Within the confines of specific systems, there is a great deal that can be done to improve retrieval once usage patterns have been properly analyzed. I do anticipate that systems will be improved slowly as the creators of those systems get a solid sense of what people are trying to do in the online environment and how they are approaching the various steps in the process.</p> <p>Our effectiveness is limited by our factory-style approach towards search [one-size-fits-all]. Information seeking behavior depends on the context of the searcher and we don't build systems that accommodate a wide variety of contexts, learning styles or formats.</p> <p>User education is a must! Intuitive design and parsing applications will only go so far. The user has to understand how the system functions (at least to some extent).</p> <p>Systems will have to be capable of recognizing instances where help might be useful (for example: the system might say to a user, "You haven't clicked on anything provided in this first page of results. Would you like help in refining your query?")</p> <p>Inconsistent human intervention. If users are to be allowed to input associated metadata for documents into a system, then you will have to build a system adequately robust to allow for inconsistent or bad behavior on the part of those users and incorporate ways for the system to retrieve content without being fully dependent on the user-generate metadata. You have to have systems that allow for inconsistencies of human behavior.</p>
--	----------	--	--	--	---

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 24 (Continued)

National Commission of Libraries and Information Science (NCLIS)	A	X				<p>Yes, there is a clear need and there is a great deal of money to be made.</p> <p>User education, online examples of great searchers given a search area (subject).</p> <p>Limited knowledge of what to expect from the system. Use more online examples.</p>
Southeastern Library Network (SOLINET)	A				X	<p>Yes, but the accuracy or relevancy of the information will not change.</p> <p>Educate the user.</p> <p>Limited knowledge which leads to being overwhelmed.</p>

SYSTEM IMPROVEMENTS AND RETRIEVAL EFFECTIVENESS

Question # 13, 14, 15, & 16

OTHER LIBRARIES RESPONSES

Table # 25

Catholic University of America	A			X		<p>Better controlled vocab., misspelling corrections, lots of searchable field limits.</p> <p>No. We are subject to too many market pressures—making our dbs like Amazon or Google.</p> <p>Education, clearly and simply written help screens.</p>
Senate Library	A		X			<p>Better relevance ranking algorithms need to be developed to enable better precision in full-text searches.</p> <p>Lack of precision in results algorithms; lack of (affordable) software to automatically categorize incoming content in databases; lack of customization, individual taxonomies and controlled vocabularies. LCSH is not a “one size fits all” controlled vocabulary.</p> <p>Get some training from an information professional on how to construct better searches; doing some digging on the database to see how the content is organized and what thesauri or metadata is used on the database, and then use those terms in conjunction with full-text searching.</p> <p>The refusal of users to consult librarians and other informational professionals is maddening. That barrier can be self-generated, or perhaps the user has had unpleasant experiences with librarians. Lousy database design—unhelpful help screens and “term not found” notices, with no mechanism for bumping people back to the original search page, or no mechanism for suggesting other terms to use if the ones they use aren’t in the database.</p>

FUTURE ROLE OF CATALOGERS AND INDEXERS

Question # 25 & 26

CENDI MEMBER AGENCIES RESPONSES

Table # 26

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p># 25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?</p> <p># 26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?</p>
Defense Technical Information Center, (DTIC)	A				X	<p># 25. Their role can be minimized! If catalogers and indexers are providing a quality product, then they are enhancing searching.</p> <p># 26. DTIC indexing, not sure it is helpful! Machine aided indexing at least provides consistency. Full text with cataloging of DTIC data! Addition of metadata is good! Catalogers and indexers must maintain a high level of quality to output!</p>
Defense Technical Information Center, (DTIC)	B			X		<p># 25. I don't believe the role of catalogers is minimized! You still need descriptive metadata.</p> <p># 26. Possibly indexers are less necessary, at least on the terms they come up with that are already in the text. Where indexers are still needed is in coming up with terms not in the actual text, synonyms or a concept talked around but not mentioned. It would also be nice to add terms later for new names for concepts and changes in author name.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 26 (Continued)

Defense Technical Information Center, (DTIC)	C	X				<p># 25. Don't believe the role is minimized. The role needs to be automated.</p> <p># 26. It is always important to have human intervention to improve the quality of your results.</p>
Defense Technical Information Center, (DTIC)	D	X				<p># 25. I believe they would play a lesser but still important role. At DTIC, the descriptive cataloging associated with the classification and any limitations on secondary distribution to the document will always be important. I believe human catalogers will play a lesser role with respect to subject cataloging.</p> <p># 26. Machines will ultimately be able to suggest all metadata, but there will a few pieces of metadata that you will always want human review to ensure that it is accurate. For example, at DTIC, I don't see humans not reviewing the metadata associated with the classification and any limitations on secondary distribution to the document anytime soon.</p>
Defense Technical Information Center, (DTIC)	E		X			<p># 25. In the world of Google, yes! In our world, no! Cataloger and indexers are even more important as the size of our collection increases; we need good catalogers to ensure accurate input, so that documents are accessible to searchers. We also need good catalogers to improve descriptors and identifiers for effective research and retrieval.</p> <p># 26. Yes, absolutely! There has to be human intervention to ensure good quality control.</p>
Defense Technical Information Center, (DTIC)	F	X	X			<p># 25. There is the need for both! Their expertise is necessary to get to the relevant documents.</p> <p># 26. Yes! What you get out of the system is only as good as the input! Human intervention is therefore critical!</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 26 (Continued)

Defense Technical Information Center, (DTIC)	G			X		# 25. The failures of full-text searching show the need of catalogers and indexers. That is why the web now uses metatags, to try to get people to catalog their own works. Organizations that have internal full-text databases often require employees to enter metadata for their own works.
Government Printing Office (GPO)	A			X		# 25. It could be, but it should not be. Even if metadata is not used to enhance the search itself, metadata is essential to the management of the full-text data over time. # 26. Yes, even the best automated metadata indexing requires management and the introspective review necessary to keep the index current and useful.
Government Printing Office (GPO)	B	X	X			# 25.No. Their role is and remains increased in the future # 26. Given the current state of technology yes, but I expect this to change in the future.
Library of Congress	A	X				#25. In an ideal world, catalogers and indexers working together with developers could make a better system. Most catalogers see their work as an art form they are not connecting people to information. #26. I would like to know!
Library of Congress	B	X	X			#25. NO! #26.YES!!!!!!

CENDI MEMBER AGENCIES RESPONSES

Table # 26 (Continued)

Library of Congress	C	X	X			<p>#25. Unfortunately yes.</p> <p>#26. Absolutely!</p>
NASA Scientific and Technical Information Program	A			X		<p># 25. Yes.</p> <p># 26. Yes, though it may be very limited, and more in the quality control area.</p>
National Agricultural Library (NAL)	A			X		<p># 25. Not at all</p> <p>Yes, especially if multiple languages and data types are to be searched.</p>
National Archives and Records Administration (NARA)	A	X	X			<p># 25. It depends. The search results retrieved can be overwhelming to users in a full-text system if the data set is very large and/or contains documents with a narrowly focused topical scope. In those cases, indexers and catalogers play an important role in providing quality metadata. Most effective systems rely on a combination of both full-text keyword and metadata searching, so I do not think that the role of catalogers and indexers is minimized in real life. If you have a database for which full-text searching is the only form of access and no catalogers or indexers are hired to encode any metadata, then, of course, their role is nonexistent.</p> <p># 26. Yes. As an information professional, I feel that metadata indexing (done by people) is important to improving the quality of search results for experienced and “power” users. Topical experts who do descriptive work and indexing work add valuable content to a database. Also, with a little</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 26 (Continued)

National Archives and Records Administration (NARA)	A (Cont.)	X	X			user education, less experienced users can improve their search experiences and results by learning to user metadata and controlled vocabulary search strategies.
National Archives and Records Administration (NARA)	B	X	X			# 25. No! Catalogers and indexers still should play a central role in highlighting the key concepts, topics, names, places, etc. that are found in the document or record. Full-text searching will not highlight the pertinent information for the end user.
National Library of Medicine (NLM) NIH	A	X	X			<p># 25. No. The availability of full text allows for more access to content, but it can also lead to information overload. Searches can yield the maximum number of hits, but not give the user the desired results in an easily understandable, digestible format. Metadata also can easily be standardized, with targeted points of entry to a document, where the number of ways of saying essentially the same thing in a piece of text is infinite. With full text searching, there are more opportunities for humans to apply metadata to documents and improve search results.</p> <p># 26. Yes. Right now, I am working on a project to use an automated tool to apply metadata to text. The tool does very well in including the proper terms when the text is abundant, but it also applies indexing terms that have nothing to do with the meaning of the text, because the tool cannot understand exactly what sentences mean. It can look at specific words and phrases, but it can't understand the difference between sentences that a human being grasps without thinking. (I think of an example of two very different sentences to illustrate my point: "After eating, Julia cut the cake"</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 26 (Continued)

National Library of Medicine (NLM) NIH	A (Cont.)	X	X			and “After eating Julia, cut the cake” have very different meanings and take place either in the past or the future, and these differences are the result of the judicious use of a comma.) The tool also fails to apply very appropriate terms when the text does not explicitly mention something that the document is very much about. Perhaps someday, human intervention will not be necessary, but so far, irrelevant terms still need to be removed and relevant terms still need to be added.
Office of Scientific and Technical Information (OSTI) DOE	A				X	# 25. Yes and will ultimately be eliminated as non-essential expenditures. # 26. Only for those classes of information numeric data, images, software, charts, audio and multimedia files!
Office of Scientific and Technical Information (OSTI) DOE	B	X				# 25. Catalogers and indexers were once essential. Now they are helpful, but not essential. # 26. If you can afford to have human created metadata, it will enable better searching. But, increasingly, human created metadata is becoming unaffordable.
USGS Biological Resources Division (Dept. of Interior)	A			X		# 25. No, I think it is still a needed support. If automation can help their process, that improves the overall process significantly. Probably a targeted approach as to what best an organization should catalog, versus everything for all users, makes sense to deal with this constant conflict between do we catalog or not. If there resources are available, clear improvements in search results can be obtained, over full-text, if high quality, skilled, and domain expert catalogers exist.

CENDI MEMBER AGENCIES RESPONSES

Table # 26 (Continued)

USGS Biological Resources Division (Dept. of Interior)	A (Cont.)			X	<p># 26. Yes, as stated above. Especially, if an organization has a number of important/special interest documents, I'm not sure how else you highlight, through weighting, this content. Metadata also is very valuable in delivering customized results and different views of content to users. The human knowledge of how best to organize the content, incorporation of user needs, and review of generate metadata is all still needed within information organizations. This may change in the future, but I wouldn't anticipate this change in the next 5 years. The need to cataloging and the processes used hasn't changed that significantly in the last 20 years. Tools are better, some terms can be generated, but classification schemes/vocabularies are all need to improve these future efforts and metadata is key for achieving this.</p>
---	----------------------	--	--	----------	---

FUTURE ROLE OF CATALOGERS AND INDEXERS

Question # 25 & 26

DOD Organizations and DOD Contractors

Table # 27

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p># 25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?</p> <p># 26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?</p>
Air Force Research Laboratory WPAFB	A		X			<p># 25. Yes</p> <p># 26. Yes – the best quality databases – e.g. – WorldCat, or Engineering Village, or DTIC – demonstrate that.</p>
Chemical and Biological Information Analysis Center (CBIAC)	A			X		<p># 25. I would hope not.</p> <p># 26. Definitely.</p>
Chemical and Biological Information Analysis Center (CBIAC)	B	X				<p># 25 & 26. No Response!</p>

DOD Organizations and DOD Contractors

Table # 27 (Continued)

Johns Hopkins University, Applied Physics Laboratory	A		X			<p># 25. Should not be! People do not understand the role of librarians! There is the perception that their role is minimized with the advent of full text searching.</p> <p># 26. There is still a need for human intervention!</p>
Johns Hopkins University, Applied Physics Laboratory	B		X			<p># 25. Metadata! Time! Better recall! The data may not be available in full text search engines!</p> <p># 26. Yes!</p>
Johns Hopkins University, Applied Physics Laboratory	C			X		<p># 25. No! Need to use all available tools!</p> <p># 26. Yes! Humans first, then machine next! The final decision should be made by human!</p>
Lackland Air Force Base	A				X	<p># 25. No, not at all! Users still will not know the importance of using synonyms, Boolean techniques or how to tweak their results to increase or decrease the quantity, relevance or accuracy</p> <p>Most definitely. Computers are great, but natural word syntax, language idioms and slang greatly affecting search capabilities. Only the human mind can make the necessary distinctions.</p> <p>#26. No, I think you will always need catalogers/indexers to get the correct metadata into the file. They are trained to do this kind of work whereas the author wouldn't know what to put where.</p>

DOD Organizations and DOD Contractors

Table # 27 (Continued)

MITRE Corporation	A	X				<p>#25. Again, depends on the application and how people will be looking for information. Research over many years has shown the inconsistency of human indexing. Basically, the indexing is one person's view of the content, which is unlikely to help a range of people looking for that information that may have different information needs and understanding of the content.</p> <p>#26. Again, I don't believe that metadata indexing does improve search results except for bibliographic metadata such as data, title, and author.</p>
MITRE Corporation	B	X	X			<p>#25. Don't know, but I suspect it has been!</p> <p>#26. Yes!</p>
Naval Research Laboratory (NRL)	A			X		<p># 25. Yes</p> <p># 26. Absolutely yes. Even electronic metadata creation cannot achieve the complex analysis that sometimes is only evident to an expert metadata creator.</p>
Pentagon Library	A	X				<p># 25. Yes</p> <p># 26. Yes – the best quality databases – e.g. – WorldCat, or Engineering Village, or DTIC – demonstrate that.</p>

DOD Organizations and DOD Contractors

Table # 27 (Continued)

US Army Library Picatinny Arsenal, NJ	A	X	X			<p># 25. Seams to be heading that way!</p> <p># 26. Yes! Machine will never be able to do everything!</p>
Redstone Scientific Information Center (RSIC)	A				X	<p># 25. Yes</p> <p># 26. It depends upon the volume of data and uniqueness of content in each result.</p>
Redstone Scientific Information Center (RSIC)	B	X				<p># 25 & 26. No Response!</p>

FUTURE ROLE OF CATALOGERS AND INDEXERS

Question # 25 & 26

UNIVERSITY PROFESSORS RESPONSES

Table # 28

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p># 25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?</p> <p># 26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?</p>
Old Dominion University	A	X	X			<p># 25. No</p> <p># 26. No</p>
Old Dominion University	B				X	<p># 25. If the only use of the catalog was to support searching, then "yes" by definition. If the catalog plays other roles (e.g., collection management) then, "no".</p> <p># 26. No. And I'm not convinced there ever was. I do believe there is a great need for human intervention in acquiring or creating the metadata. The subsequent indexing of it should be the easy part</p>
Syracuse University	A		X			# 25 & 26. No Comment!

UNIVERSITY PROFESSORS RESPONSES

Table # 28 (Continued)

Syracuse University	B	X	X			# 25. I'd say their role is shifted rather than minimized. Metadata may be generated automatically, but the data generated still need human catalogers to verify. Much of their time will be spent on this task. Maintaining controlled vocabularies and create new ones and mapping between keywords and controlled vocabularies are all going to occupy more of their time than before. Definitely.
San Jose University	A		X			# 25. No. # 26. Definitely.
University of North Carolina	A			X		# 25. No, the catalogers of the future will not be assigning terms to objects; they will be tuning data mining algorithms. We need them like we needed catalogers in the past.

FUTURE ROLE OF CATALOGERS AND INDEXERS

Question # 25 & 26

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 29

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p># 25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?</p> <p># 26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?</p>
Access Innovation Inc.	A				X	<p># 25. It has been yes but then I don't think full text alone is the answer</p> <p># 26. Yes, I think a lot of the terms can be gathered automatically but they need to be humanly reviewed for the 15 = 30 % that is incorrectly presented for each information object.</p>
Information International Associates Inc.	A	X	X			<p># 25. If full text becomes the form of search, of course. I think the question is more of balance and how much manual intervention is appropriate. Clearly as machine tools continue to improve and the cost benefit changes, it has and will continue to minimize the role of catalogers and indexers. I think there needs to be new assessments of how these things work together. Increasingly the machine becomes more the doer and the human becomes the quality controller. Also the subject matter expert is important to ensure the search algorithms stay honest.</p> <p># 26. Yes, I don't think we've made machines smart enough yet not to</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 29 (Continued)

Information International Associates Inc.	A (Cont.)	X	X			have manual oversight and intervention on search. But things have changed significantly and roles have to be reassessed to deal with the realities that already exist, yet continue to find the role for quality that machines cannot yet fathom. The human defines what is wanted and discerns whether the results answer the questions.
Information International Associates Inc.	B			X		# 25. Yes, the tradition role is, but the mistake that organizations make is to think that they don't need their skills. The fact is that to do it right there needs to be attention to the knowledge bases that sit behind the search engines. The catalogers and indexers may do less actual metadata creation, but they do more knowledge base and rules development. # 26. Yes, see my answer above.
National Federation of Abstracting & Information Services (NFAIS)	A				X	# 25. No. I think that intelligent indexing by those with knowledge of the field is a highly desirable if costly value. Users # 26. Yes, as long as language is ambivalent in its usage. Another set of eyes and another brain in assessing and evaluating the content is always desirable.
National Commission of Libraries and Information Science (NCLIS)	A	X				# 25. To a degree # 26. Yes
Southeastern Library Network (SOLINET)	A				X	# 25. No. # 26. Certainly.

FUTURE ROLE OF CATALOGERS AND INDEXERS

Question # 25 & 26

OTHER LIBRARIES RESPONSES

Table # 30

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p># 25 Do you believe that the role of catalogers and indexers is minimized by using full-text searching?</p> <p># 26. Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?</p>
Catholic University of America	A			X		<p># 25. No Response!</p> <p># 26. Of Course!</p>
Senate Library	A		X			<p>#25. If catalogers are not constructing taxonomies or thesauri, or assigning metadata terms to be used with the databases, it is very easy for administrators to say that catalogers and indexers are no longer necessary.</p> <p>#26. Absolutely!</p>

IMPROVING SEARCH RESULTS...METADATA AND FULL TEXT SEARCHING

Questions # 19, 20 &21

CENDI MEMBER AGENCIES RESPONSES

Table # 31

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p>Improving Search Results...Metadata and Full Text Searching</p> <p># 19. What role should metadata play in improving search results?</p> <p># 20. Does full-text searching eliminate the requirements to construct metadata?</p> <p># 21. Can full-text search be used to effectively augment metadata?</p>
Defense Technical Information Center, (DTIC)	A				X	<p># 19. Metadata can greatly improve the information we want to identify.</p> <p># 20. It depends on what you are trying to do! Labor cost! It is intensive!</p> <p># 21. Yes!</p>
Defense Technical Information Center, (DTIC)	B			X		<p># 19. A lot! Helps in narrowing the search! For example, when searching for a study by a specific organization, the meta tags will enhance your results!</p> <p># 20. No! I don't believe so! You need both the descriptive metadata, such as author, title etc., and the subject metadata for synonyms to the words in the article.</p> <p># 21. Yes! Will get some false hits too!</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 31 (Continued)

Defense Technical Information Center, (DTIC)	C	X				<p># 19. Meta tagging should complement full text searching.</p> <p># 20. No. Full text searching demands adding meta tagging to make the searching useful.</p> <p>#21. They should complement each other.</p>
Defense Technical Information Center, (DTIC)	D	X				<p># 19. Metadata can play a role in the categorization. For example, categorization tools can use the metadata to augment full text searching by placing the results in categories based on the metadata, e.g. all reports by a particular corporate author are placed in one bucket, etc...</p> <p># 20. No. Still need metadata for classification / limitations on the document, etc.</p> <p>#21. Yes! Though I believe the more relevant statement is that metadata can be used to augment full text searching.</p>
Defense Technical Information Center, (DTIC)	E		X			<p># 19. Data needs to be input correctly and accurately. There needs to be consistency!</p> <p># 20. No! There still has to be metadata! There is a difference between digitization and preservation! There is the need to preserve the metadata or descriptions near the files described.</p> <p># 21. Yes!</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 31 (Continued)

Defense Technical Information Center, (DTIC)	F	X	X		<p># 19. Controlled Vocabulary is important! Also, by providing browse features, e.g., lists of authors to determine how the authors name appear in the documents to be retrieved.</p> <p># 20. No!</p> <p># 21. Yes! The two work together well!</p>
Defense Technical Information Center, (DTIC)	G			X	<p># 19. The same role it currently plays in most databases. It is used in field searching so you can find reports by authors rather than about people, so you can find reports with a title rather than those that cite a title, so you can find reports by the Army rather than those that just mention ‘army’, etc.</p> <p># 20. No. That is why we now have meta tags in html and is one of the reasons why xml has been developed.</p> <p># 21. Sometimes, if the metadata quality is poor. More often than not, it just expands one’s search results to include many marginally relevant documents.</p>
Government Printing Office (GPO)	A			X	<p># 19. In addition to providing tools to get at specific things in text, it affords a structure for the development of the consistency necessary for a user to confidently construct quality searches.</p> <p># 20. Absolutely not, it is an opportunity to enrich the text.</p> <p># 21. I believe that full text search information can be used to augment and improve metadata by providing insight into how users view the information.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 31 (Continued)

Government Printing Office (GPO)	B	X	X			<p># 19. A very critical role. It is highly underutilized and is too often thought of as simply a spamming method of achieving higher relevancy rankings.</p> <p># 20. No.</p> <p># 21. Yes.</p>
Library of Congress	A	X				<p># 19. Helpful in getting searchers to the information they need, but must understand how the system works!</p> <p># 20. Probably not!</p> <p># 21. Yes!</p>
Library of Congress	B	X	X			<p># 19. The more metadata the better!</p> <p># 20. No! No! No!</p> <p># 21. YES</p>
Library of Congress	C	X	X			<p># 19. Role? Perhaps the question needs rephrasing: To what degree should metadata improve results? Again, it depends on the quality of the metadata. "Garbage in, garbage out."</p> <p># 20. Does it or should it? Yes it does, no it should not.</p> <p># 21. Yes. And vice versa.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 31 (Continued)

NASA Scientific and Technical Information Program	A			X		<p># 19. It could improve results ---- only if the metadata itself improves and is more widely available for all data.</p> <p># 20. No. Depends upon other factors. See also prior comments.</p> <p># 21. Yes, since a mix may be the optimal approach.</p>
National Agricultural Library (NAL)	A			X		<p># 19. Must be there in the content, may also be used for advanced searching. For the people who like browsing, controlled vocabulary a must, and even if the browsers are a minority, they can be an important minority.</p> <p># 20. No, esp. if multiple the text includes numeric tables, charts, etc. multiple languages and alphabets.</p> <p># 21. Yes</p>
National Archives and Records Administration (NARA)	A	X	X			<p># 19. If the search system bases its relevancy rankings in part on whether (and where) the search query terms appear in the metadata, this can be a valuable way to improve search results.</p> <p># 20. As I mentioned above in my response to #18, visible metadata and controlled vocabulary can help researchers retrieve all of the results that include that precise term. This can increase the precision of the database and improve users' search results.</p> <p>In general, I don't think so. More specifically, I think it depends on how valuable the data is and how precise retrieval needs to be to sufficiently serve internal business needs and/or the public's information needs. (Constructing metadata can be a costly expense, as it might call for a serious commitment of staff time or outsourcing dollars.) If the</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 31 (Continued)

National Archives and Records Administration (NARA)	A (Cont.)	X	X		<p>information is at all valuable and important, I would venture that some sort of metadata or controlled vocabulary indexing would be necessary to ensure accurate retrieval. Full-text searching would not be enough if the users (internal and external to the organization) frequently needed to retrieve <i>all</i> relevant material from the database. For example, although undergraduate students only need to retrieve enough relevant information to write their term papers, during “discovery” for litigation purposes, lawyers and staff are required to produce all relevant documents as evidence.</p> <p># 21. Yes, the two are often used together successfully to produce more precise search results for users at various levels.</p>
National Archives and Records Administration (NARA)	B	X	X		<p># 19. Controlled vocabulary or thesauri</p> <p># 20. No. It is a compliment to full text searching, not a replacement.</p> <p># 21. Yes.</p>
National Library of Medicine (NLM) NIH	A	X	X		<p># 19. Metadata makes search results more relevant. Metadata should be weighted more heavily than the text of the document. It relates directly to the meaning, significance, timeliness, creator, and publisher of the document. The text may or may not mention these pieces of data.</p> <p># 20. Absolutely not! Sometimes the text mentions the context and subjects of the document, but not always.</p> <p># 21. Yes. We should use all the information at our disposal to improve search results. If we have the full text, there is no reason to ignore it.</p>

CENDI MEMBER AGENCIES RESPONSES

Table # 31 (Continued)

Office of Scientific and Technical Information (OSTI) DOE	A				X	<p># 19. Very little! Unless it describes numeric data, images, software, charts, audio and multimedia files.</p> <p># 20. It depends on the nature of the document.</p> <p># 21. I think the question is stated in reverse order. Metadata has very limited use except for numeric data, images, software, charts, audio and multimedia files.</p>
Office of Scientific and Technical Information (OSTI) DOE	B	X				<p># 19. The way you search a database depends on the richness of the database. If you have well constructed metadata, then you use it. If you don't have well constructed metadata, you can still search via the full text.</p> <p># 20. Full text searching makes it possible to search without metadata. Metadata helps enable searching, but, increasingly, creating metadata is becoming unaffordable.</p> <p># 21. Yes.</p>
USGS Biological Resources Division (Dept. of Interior)	A				X	<p># 19. I mentioned this in previous questions. Items such as: weighting of results, sub-setting information display to a user, suggesting like or additional items, quality ranking/rating, helping a user visualize the information repository, narrowing results based on some criteria provided by the user (for instance if someone puts in a Base Name, you know they are probably look for reports related to a certain base or organization). Inference such as this can also be built upon metadata and aid users.</p> <p># 20. No!</p> <p># 21. Yes it can. If common keywords can effectively be generated to accurately represent a document, if no metadata exists, and/or no domain experts exist to create the metadata, full text can augment the metadata.</p>

IMPROVING SEARCH RESULTS...METADATA AND FULL TEXT SEARCHING

Questions # 19, 20 &21

DOD Organizations and DOD Contractors

Table # 32

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p>Improving Search Results...Metadata and Full Text Searching</p> <p># 19. What role should metadata play in improving search results?</p> <p># 20. Does full-text searching eliminate the requirements to construct metadata?</p> <p># 21. Can full-text search be used to effectively augment metadata?</p>
Air Force Research Laboratory WPAFB	A		X			<p>No Comment</p> <p>No!</p> <p>Yes!</p>
Chemical and Biological Information Analysis Center (CBIAC)	A			X		<p>No Response!</p> <p>No!</p> <p>Yes!</p>
Chemical and Biological Information Analysis Center (CBIAC)	B			X		<p>No Response!</p> <p>No!</p> <p>Yes!</p>

DOD Organizations and DOD Contractors

Table # 32 (Continued)

Johns Hopkins University, Applied Physics Laboratory	A		X		<p># 19. Metadata should automatically map to content! What terms are important to search results!</p> <p># 20. No!</p> <p># 21. Yes! Product names!</p>
Johns Hopkins University, Applied Physics Laboratory	B		X		<p># 19. Useful in improving relevancy (Recall)! Limiters, for example, the first 500 hits!</p> <p># 20. No! Not if the working world must go on! There only so many hours in the day!</p> <p># 21. Yes! After one has exhausted metadata searching, full text searching is a second choice!</p>
Johns Hopkins University, Applied Physics Laboratory	C			X	<p># 19. Increase depending on the database. Metadata not necessary with photographic databases.</p> <p># 20. No!</p> <p># 21. Yes! They augment each other! There are drawbacks in standard full text searching. Words are full text! This does not take care of homogeny.</p>
Lackland Air Force Base	A			X	<p># 19. Metadata should assist in improving search results, not act as a barrier to effective searching</p> <p># 20. No, not at all. If for some reason results are less than expected, metadata might be the only other way to extract the data.</p> <p># 21. Yes</p>

DOD Organizations and DOD Contractors

Table # 32 (Continued)

MITRE Corporation	A	X				<p>#19. Good metadata is not widely supplied with government information and searching by metadata requires you to know the appropriate government jargon to match.</p> <p>#20. No. Metadata is essential for bibliographic information—author, title, date, etc.</p> <p>#21. ...unless you are searching for a specific author or title or date, full-text search is essential.</p>
MITRE Corporation	B	X	X			<p>#19. It is good in refining search results! The biggest problem is getting people to know and learn so as to improve search results!</p> <p>#20. In Google search metadata searching is not needed! With the searching of pharmaceutical databases for example, an 80/20 precession/recall does not cut it! For specific collections, metadata is needed!</p> <p>#21. Yes!</p>
Naval Research Laboratory (NRL)	A			X		<p>#19. I think it is important as a primary level. The basic and most critical elements are defined.</p> <p>#20. Absolutely not.</p> <p>#21. Perhaps.</p>
Pentagon Library	A	X				<p>#19. Improve the accuracy!</p> <p>#20. No! Make it more important!</p> <p>#21. Yes!</p>

DOD Organizations and DOD Contractors

Table # 32 (Continued)

US Army Library Picatinny Arsenal, NJ	A	X	X			<p>#19. Every document in a database should have metadata/bibliographic data for search retrieval.</p> <p>#20. Absolutely not!</p> <p>#21. Yes, first find results based on controlled vocabulary/bibliographic data/metadata, then search by full text for specific items.</p>
Redstone Scientific Information Center (RSIC)	A				X	<p>#19. No Response!</p> <p>#20. It depends on the content of the database. For example, more scientific information is easier than social science databases.</p> <p>#21. Using vague search terms can retrieve too much information or low relevancy.</p>
Redstone Scientific Information Center (RSIC)	B	X				No Response!

IMPROVING SEARCH RESULTS...METADATA AND FULL TEXT SEARCHING

Questions # 19, 20 &21

UNIVERSITY PROFESSORS RESPONSES

Table # 33

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						Improving Search Results...Metadata and Full Text Searching # 19. What role should metadata play in improving search results? # 20. Does full-text searching eliminate the requirements to construct metadata? # 21. Can full-text search be used to effectively augment metadata?
Old Dominion University	A	X	X			# 19. Significant # 20. No # 21. Yes
Old Dominion University	B				X	# 19. Well, if it improves the result that is its role. If it doesn't, if someone has a full-text search engine that do just as well, then metadata has no role in improving search results. Your question suggests an idea that metadata is an end in and of itself. I see it instead as a means. # 20. Metadata may often have other non-searching roles to play. # 21. To augment the metadata itself? Certainly. For example, given a set of metadata for some document that is missing the name of the author or that has only a partial name, one might do a Google search for the document title and come up with a page describing the author, including the author's full name.
Syracuse University	A		X			No Response!

UNIVERSITY PROFESSORS RESPONSES

Table # 33 (Continued)

Syracuse University	B	X	X			<p># 19. Offering fielded search, organizing search results by categories, and displaying results in a consistent look.</p> <p># 20. Definitely not.</p> <p># 21. Could be.</p>
San Jose University	A		X			<p># 19. Not sure about the question.</p> <p># 20. Not at all.</p> <p># 21. Not at all.</p>
University of North Carolina	A			X		<p># 19. Metadata is one of many sources of evidence for searchers---it is too expensive to produce manually for any but the most crucial corpuses so we will learn to live with automatically generated metadata.</p> <p># 20. No, generate as much as you can automatically (e.g., cameras with time and spatial codes automatic for free give good power for search)</p> <p># 21. Of course---that is what people want anyway---not metadata---metadata is a means to the ends of full document</p>

IMPROVING SEARCH RESULTS...METADATA AND FULL TEXT SEARCHING

Questions # 19, 20 &21

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 34

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p>Improving Search Results...Metadata and Full Text Searching</p> <p># 19. What role should metadata play in improving search results?</p> <p># 20. Does full-text searching eliminate the requirements to construct metadata?</p> <p># 21. Can full-text search be used to effectively augment metadata?</p>
Access Innovation Inc.	A				X	<p># 19. A BIG ONE!! Control of the terms in use and the way they are applied is crucial. For all the time we have spent on this to date we have not deployed the indexing part very well at all.</p> <p># 20. No, it makes it more important due to all the false drops from using the same term in different meanings and as pictures speech.</p> <p># 21. Yes – I think that was the original idea.</p>
Information International Associates Inc.	A	X	X			<p># 19. Structured controlled vocabulary to get to concepts. Limiting search results like if you want to know where a thing was published vs. where the investigation focused on, fielded metadata search can be a discriminator.</p> <p># 20. It depends on users, context and objective of system. It increasing makes it less useful for general, fast quick and dirty searching as evidenced by the big search engines. But if one needs to get statistics and structure of content from the DB, then metadata is still critical.</p> <p># 21. Absolutely!</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 34 (Continued)

Information International Associates Inc.	B			X	<p># 19. I think I've addressed this in the last question. I hope.</p> <p># 20. No, as I said, I think the best way is both. Also, metadata is the way to provide links between documents. It is also the way to add information that does not appear or is not readily discernable from the document itself. For example, document type.</p> <p># 21. I would say that they would augment each other.</p>
National Federation of Abstracting & Information Services (NFAIS)	A			X	<p># 19. One should not be limited solely to metadata for searching purposes but one should ensure that metadata is always available for that purpose. That means it should be consistently reliable and made searchable according to the need of the user.</p> <p># 20. No, it is imperative that we have <u>both</u> in place so that users with different through processes and learning approaches can be successful in their information seeking. That's why it is so crucial that we learn from user data, drawing from the broadest possible variety of real world queries and content formats.</p> <p>#21. It would be lovely to be able to turn it off and on as need dictated.</p>
National Commission of Libraries and Information Science (NCLIS)	A	X			<p># 19. Have vendors use basic sets of metadata so people could learn that a given set of metadata will be there and can be searched.</p> <p># 20. No!</p> <p>#21. Yes!</p>

INFORMATION SCIENCE ORGANIZATIONS RESPONSES

Table # 34 (Continued)

Southeastern Library Network (SOLINET)	A				X	# 19. Cross-referencing. # 20. No! # 21. No!
---	---	--	--	--	---	--

IMPROVING SEARCH RESULTS...METADATA AND FULL TEXT SEARCHING

Questions # 19, 20 &21

OTHER LIBRARIES RESPONSES

Table # 35

	PARTICIPANT	F U L L T E X T	M E T A D A T A	O T H E R	N O P R E F	COMMENTS
						<p>Improving Search Results...Metadata and Full Text Searching</p> <p># 19. What role should metadata play in improving search results?</p> <p># 20. Does full-text searching eliminate the requirements to construct metadata?</p> <p># 21. Can full-text search be used to effectively augment metadata?</p>
Catholic University of America	A			X		<p>#19. Don't know. I'm still confused by your use of term metadata.</p> <p>#20. If use of metadata –whatever that is--results in a more precise search, then I'm for constructing metadata.</p> <p># 21. Well, whenever you do a full text search you will increase retrieval, so it depends on your goal.</p>
Senate Library	A		X			<p># 19. It's vital!</p> <p># 20. Absolutely not. Searching “automobiles” won't find documents that are about “cars” unless a taxonomy/thesaurus schema is running in the background to guide people.</p> <p># 21. Oh, of course. I use full-text searching all the time (on Google, that's all there is).</p>

APPENDIX: B

INTERVIEW QUESTIONNAIRE:

SEARCHING METHODOLOGY STUDY

September 2007

Interview Questions and Statements for Comments

1

Scholars often refer to full text searching as searching devoid of controlled vocabulary, taxonomies, subject classification, metadata, etc., when in fact; most full-text databases often incorporate some form of classification, structure, complex search algorithms, bibliographic fields and abstracts.

Please Comment.

2

Early web search engines relied on Boolean methodology in meeting full-text searching needs. Later web programmers gradually began applying metadata, taxonomies, and algorithms. We now find bibliographic and full-text information combined. Is the real issue therefore, what recipe of metadata, taxonomies, and algorithm to apply.

Please Comment.

3

In discussing database comparison, the issue should be about content retrieval rather than about search functions. The ultimate goal is find the information that a searcher is seeking.

Please Comment.

4

Generally, end users are not expert searchers; therefore their results are a function of their limited searching capability.

Please Comment.

5

End users ability to obtain relevant information from metadata searching is more limited than if full text searching is used.

Please Comment.

6

The quality of full-text searching has greatly improved due to search engines built in capabilities such as “suggested terms to the user.” On the other hand, successful search results derived from metadata searching is more a function of how well the user understands and use the available controlled vocabulary.

Please Comment.

7

In full-text searching, relevance ranking is a problem with large documents, with multiple volumes and sections, when a term is not used frequently. Such documents are assigned low relevancy ranking.

Please Comment.

8

There are limitations when using full-text searching of databases with mixed documents (multi-media and text) since there is a need to describe the document (metadata) to improve ones search results.

Please Comment.

9

Scholars often comment that if searchers had access to more accurate search systems, they would be more successful in their search results. Could it be that search systems are already “good enough,” so that a more accurate system would provide at best only marginal improvements? **Please Comment.**

10

What are some of the fundamental flaws in measuring a search engine performance, and how does one overcome these issues?

11

In the 1970's, the ideal recall and precession rate of 0.70 was considered ideal or acceptable. In light of the improvements in searching methodologies over the past 30 years, what would you consider as an acceptable recall and precession and recall value?

12

How can these levels be improved?

13

Do you anticipate any large scale improvements in retrieval effectiveness? Explain.

14

For the past 50 years or so, the challenge has been to improve the accuracy of search systems, by so doing, users will be better able to find the information that is needed. What are some of the limitations that need to be overcome for us to see more effective search systems?

15

What are some of the ways to improve user search results?

16

What are some of the barriers to a user search experience, and how can they be overcome or at least minimized?

17

What is your preferred method in searching databases for access to government information?

____ Full Text ____ Metadata ____ Other ____ No Preference

_____ Specify

18

Explain the reason for your choice?

.

19

What role should metadata play in improving search results?

20

Does full-text searching eliminate the requirements to construct metadata?

21

Can full-text search be used to effectively augment metadata?

22

What are some of the drawbacks in using full-text searching?

23

What are some of the drawbacks in using metadata searching?

24

Which searching methodology is more effective when accessing Scientific and or quantitative data, and why?

25

Do you believe that the role of catalogers and indexers is minimized by using full-text searching?

26

Do you believe that there is still a need for human intervention in metadata indexing to improve the quality of search results?

Thank you for participating in this study.

REFERENCES

- Bandos, J., and Resnick, M.L. (2004). The Effects of Semantic and Syntactic Instruction on User Performance and Satisfaction in Search User Interface Design. Proceedings of the Human Factors and Ergonomics Society 48th Annual Conference. Human Factors and Ergonomics Society.
- Bates, M.J. Information Search Tactics. *Journal of the American Society for Information Science*, 30, July 1979, 205-214.
- Bates, M.J. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review*, 13, October 1989, 407-424.
- Bates, M.J. (2003). Improving User Access to Library Catalog and Portal Information: Final Report. In: <http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf>
- Belkin, N.J., Oddy, R.N., and Brooks, H.M. ASK for Information Retrieval: Part I. Background and Theory. *The Journal of Documentation*, Vol. 38, No 2, June 1982, pp. 61-71.
- Belkin, N.J., Oddy, R.N., and Brooks, H.M. ASK for information retrieval: Part II. Results of a Design Study. *The Journal of Documentation*, Vol. 38, No 3, September 1982, pp. 145-164.
- Belkin, N.J., Croft, B. W. Information Filtering and Information Retrieval: Two Sides of the Same Coin. *Communications of the ACM*, 35 (12), December 1992, 29-38.
- Blair, D and Maron, M.E., an Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System. *Communications of the ACM*, 28 (3), March 1985, 290-299.
- Borgman, C.L. (1986a). Why are Online Catalogs Hard to Use? Lessons Learned from Information Retrieval Studies. *Journal of the American Society for Information Science*, 37 (6), 387-400.
- Borgman, C.L. (1996). Why are Online Catalogs Still Hard to Use? *Journal of the American Society for Information Science*. 47 (7): 493-503.
- Bourdoncle, Francois. Enabling Effective Human-Machine Interaction. Presented at the 12th. Search Engine Meeting, Boston, MA. April 23, 2007.
- Byrum, John D. (2005). Recommendations for Urgently needed Improvement of OPAC and the Role of the National Bibliographic Agency in Achieving it. Paper presented at the 71st IFLA General Conference and Council, Oslo, Norway, August 14-18. Available: <http://www.ifla.org/IV/ifla71/papers/124e-Byrum.pdf>.
- Calhoun, Karen. The Changing Nature of the Catalog and its Integration with other Discovery Tools. Final Report: Prepared for the Library of Congress. March 17, 2006.
- Cooper, M.D. Usage Patterns of a Web-Based Library Catalog. *Journal of the American Society for Information Science and Technology*, 52, 137-148, 2001.
- Davies, Clement. Reference Retrieval by User-Negotiated Term Frequency Ordering within a Dynamically Adjusted Notional 'Document.' *Journal of Informatics*, 2, 1978, 62-76.
- Davis, L. (2006). Improving Internet Interaction: From Theory to Practice. *Journal of the American Society for Information Science and Technology*, 57, 788-791.
- Dupuis, J. (2006). Confessions of a Science Librarian. In <http://jdupuis.blogspot.com/2006/12/my-job-in-10-years-collections-further.html>.
- Feldman, Susan. (2004). The High Cost of not Finding Information. Retrieved December 11, 2006, from <http://www.kmworld.com/Articles/ReadArticle.aspx?ArticleID=9534>.
- Gentile, Tony. (2006). A Healthy Perspective on Search Behavior. Presented at the 11th Search Engine Meeting, Boston, MA.

REFERENCES

- Gremett, P. (2006). Utilizing a User's Context to Improve Search Results. *Journal of the American Society for Information Science and Technology*, 57, 808-812.
- Gulli, A. and Signorini, A. The Indexable Web is More than 11.5 billion pages. *www* 2005, May 10-14, 2005, in <http://www.cs.uiowa.edu/~assignori/web-size/>
- Hawking, D, and Thomas, P., Server Selection Methods in Hybrid Portal Search. SIGIR'05, August 15-19, 2005, Salvador Brazil.
- Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., & Yee, K. (2002). Finding the Flow in Web Site Search. *Communications of the ACM*, 45 (9), 42-49.
- Henseler, Hans (2005). Usability over High Precision and Recall. Presented at the 10th Search Engine Meeting. Boston, MA.
- Jacso, Peter. Savvy Searching: Relevance in the Eye of the Search Software. *Online Information Review*. Vol 29, No 6, 2006. 676-682.
- Jansen, B.J. (2005). Seeking and Implementing Automated Assistance During the Search Process. *Information Processing and Management*, 41, 909-928.
- Jefferson, T.I., and Nagy, T.J. (2002). A Domain-Driven Approach to Improving Search Effectiveness in Traditional Online Catalogs. *Information and Management*, 39, 559-570.
- Jones, M., Buchanan, G., & Thimbleby, H. (2003). Improving Web Search on Small Screen Devices. *Interacting with Computers*, 15, 479-495.
- Kerchner, Marcia. (2006). A Dynamic Methodology for Improving the Search Experience. *Information Technology and Libraries*, June 2006.
- Lancaster, F.W., Ulvila, J.W., Humphrey, S.M., Smith, L.C., Allen, B., and Herner, S. (1966). Evaluation of Interactive Knowledge-Based Systems: Overview and Design for Empirical Testing. *Journal of the American Society for Information Science*, 47, 57-69.
- McKay, Andrew. (2006). The Future of Search Content Synergy. Presented at the 11th Search Engine Meeting, Boston, MA.
- Ma, Wei. (2002). A Database Selection Expert System Based on Reference Librarian's Database Selection Strategy: A Usability and Empirical Evaluation. *Journal of the American Society for Information Science and Technology*, 53(7): 567-580.
- Mann, Thomas. The Changing Nature of the Catalog and its Integration with other Discovery Tools. Final Report: Prepared for the Library of Congress by Karen Calhoun. A Critical Review. Prepared for AFSCME 2910. April 3, 2006.
- Markey, Karen. Twenty-Five Years of End-User Searching, Part 1: Research Findings. *Journal Of the American Society for Information Science and Technology*, 58 (8): 1071-1081, 2007.
- Markey, Karen. Twenty-Five Years of End-User Searching, Part 2: Future Search Directions. *Journal of the American Society for Information Science and Technology*, 58 (8): 1123-1130, 2007.
- Markhoff, John and Wyatt, Edward. "Technology; Google is Adding Major Libraries to its Database." *New York Times*, Dec. 14, 2004.
- Maybury, M. (2005). Making Search and Discovery Work. Presentation, 30, March 2005, The MITRE Corporation.
- Medeiros, Norm. (1999). Driving with Eyes Closed: the Perils of Traditional Catalogs and Cataloging in the Internet Age. *Library Computing* 18 (4): 300-305.
- Moran, Mike. (2006). Don't Just Change the Search Engine! Presented at the 11th Search Engine Meeting, Boston, MA.

REFERENCES

- Mostafa, Javed. Seeking Better Web Searches. *Scientific American*, February 2005, Vol. 292 Issue 2, p 66-73.
- Nicholson, Scott. (2004). A Conceptual Framework for the Holistic Measurement and Cumulative Evaluation of Library Services, *Journal of Documentation*, 60 (2), 2004.
- Pedersen, Jan. (2005). Internet Search Engines: Past and Future. Presented at the 10th Search Engine Meeting, Boston, MA.
- Peterson, Kim. How "Search" is Redefining the Web – and Our Lives. *The Seattle Times*, Sunday, May 1, 2005.
- Perez, Juan Carlos. What is Vertical Search? *InfoWorld*, January 18, 2006.
- Regan, Keith. Vertical Search Market Quickly Becoming Crowded. *E-Commerce Times*, April 4, 2005.
- Rose, Daniel E, and Levinson, Danny. (2004). Understanding User Goals in Web Search. WWW 2004, May 17-22, 2004, New York.
- Rappaport, A. (2002, March). Web Search Engines: Users Surprising the Expert. A Presentation to the Bay CHI Special Interest Group. Retrieved June 12, 2002, from <http://www.searchtools.com/info/slides/baychi/2002/index.html>.
- Reamy, Tom. (2006). Faceted Navigation: An Alternative to Search and Browse. Presented at the 11th Search Engine Meeting, Boston, MA.
- Resnick, M and Vaughan, M. (2006). Best Practices and Future Visions for Search User Interfaces. *Journal of the American Society for Information Science and Technology*, 57(6):781-787.
- Salton, G., Automatic Text Analysis. *Science*. 168, 3929 (April 1970), 335-343.
- Salton, G and Buckley, C. Improving Retrieval Performance by Relevance Feedback, *JASIS* 41, 1990, 288-297.
- Salton, G., Wong, A and Yang, C.S. A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 1Volume 18, Number 11, November 1975.
- Salton, G. and McGill, M.J. *Introduction to Modern Information Retrieval*, McGraw-Hill, 1983.
- Scheffer, Oliver. From Broad to Vertical Search: Retrieving Information from the Deep Web. Presented at the 12th Search Engine Meeting, Boston, MA, April 23-24, 2007.
- Silverstein, C., Henzinger, M., Marais, H., and Moricz, M. Analysis of a Very Large Web Search Engine Query Log. *SIGIR Forum*, 33 (3), 1999. Originally published as DEC Systems Research Center Technical Note, 1998.
- Sokullu, Emre. Search 2.0 – What's Next? *Read/Write Web*, December 13, 2006.
- Spink, A., Jansen, B.J., Wolfram, D., and Saracevic, T. From E-Sex to E-Commerce: Web Search Changes. *IEEE Computer*, 35 (3), 107-109, 2002.
- Sullivan, D. (2000). NPD Search and Portal Site Study. Retrieved from <http://www.searchenginewatch.com/reports/npd.htm>.
- Swanson, D.G. Searching Natural Language Text by Computer. *Science*. 132, 3434 (October 1960), 1099-1104.
- Thatcher, A. (2000). Experience and WWW Searching, or the Search for the WWW "Expert." In *Proceedings of the IEA 2000 HFES Congress*. Santa Monica, CA: Human Factors and Ergonomics Society.

REFERENCES

Tenopir, Carol. (2005). What We Know About User Behavior. Presented at the 10th Search Engine Meeting, Boston, MA.

Wouters, Paul and Vries, de Repke. Formally Citing the Web. *Journal of the American Society for Information Science and Technology*. 55, (14): 1250-1260. December 2004.